







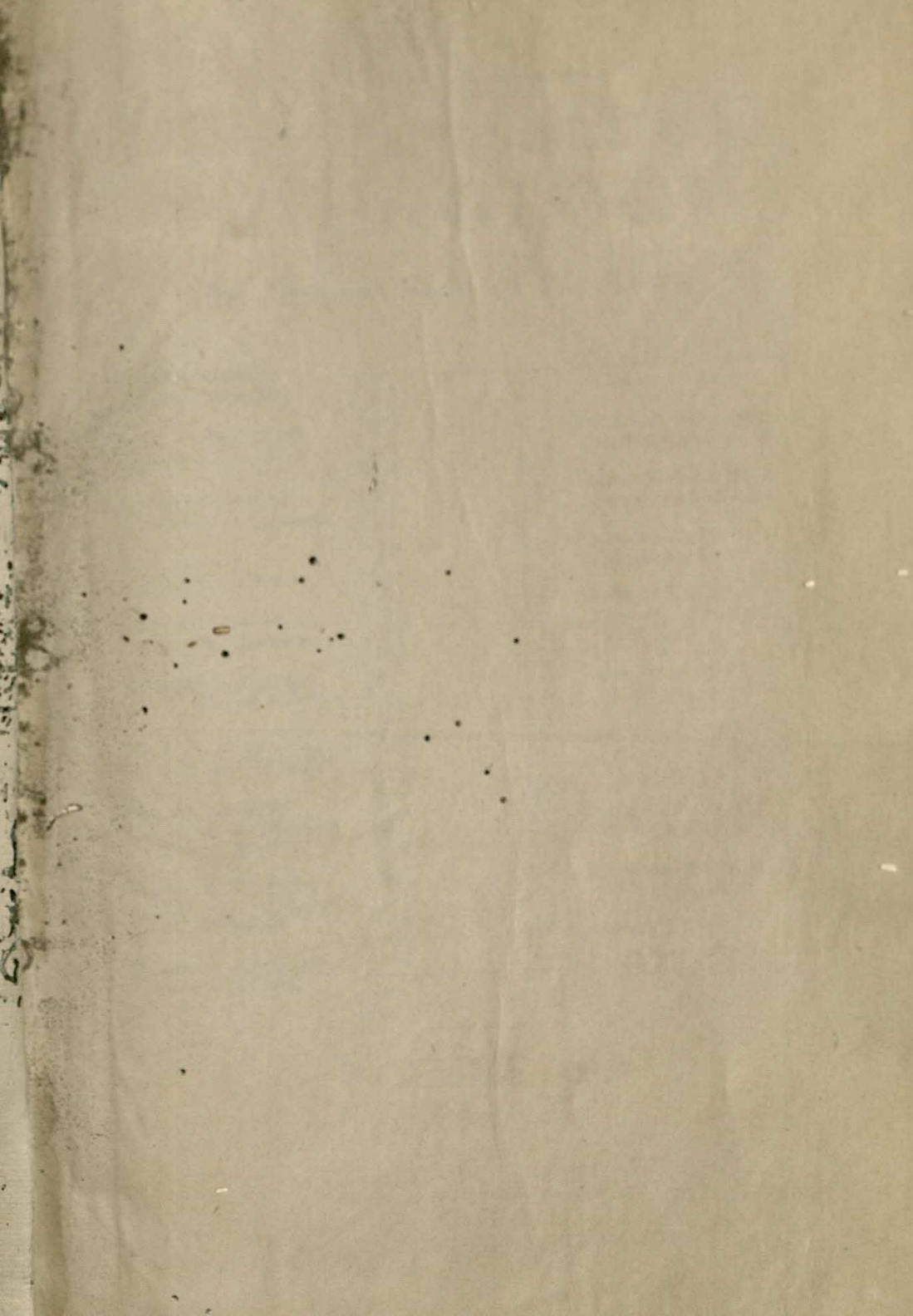


544  
9.2.77

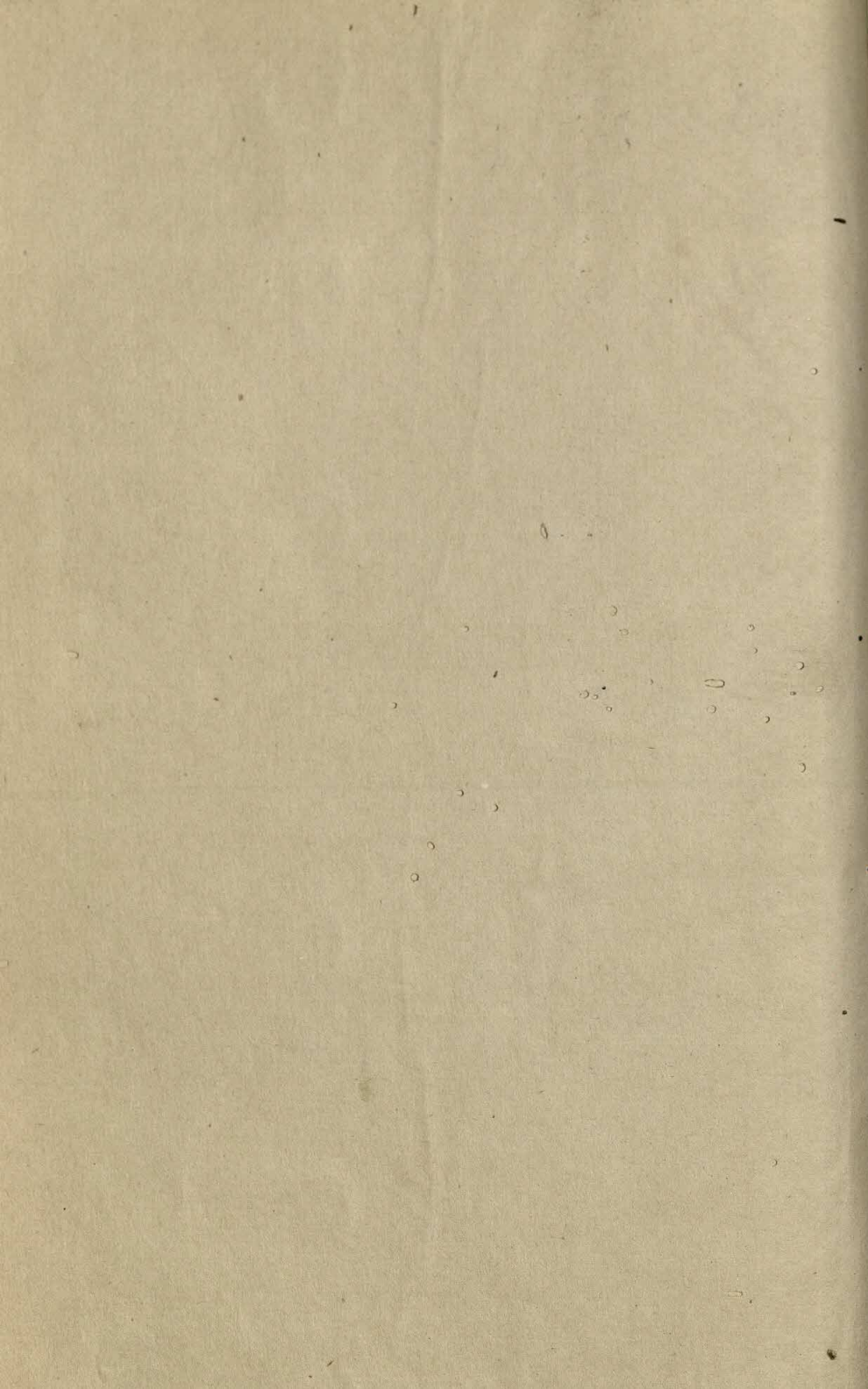














(Founded in 1935 by Carl Murchison)

# The Journal of PSYCHOLOGY

The General Field of Psychology

GORDON ALLPORT  
Harvard University

JOHN A. ANTOINETTI  
Hartford, Connecticut

ARTHUR G. BILLS  
University of Cincinnati

E. A. BOTT  
University of Toronto

LEONARD CARMICHAEL  
Smithsonian Institution

HERBERT S. CONRAD  
U. S. Office of Education

JOHN F. DASHIELL  
Florida State University

ROY M. DORCUS  
University of California  
at Los Angeles

FRANKLIN FEARING  
University of California  
at Los Angeles

FRANK A. GELDARD  
Princeton University

CLARENCE H. GRAHAM  
Columbia University

ERNEST R. HILGARD  
Stanford University

HEINRICH KLÜVER  
University of Chicago

LEONARD MEAD  
Tufts University

NORMAN L. MUNN  
University of Adelaide,  
Australia

JOHN PAUL NAFE  
Florida State University

LEE EDWARD TRAVIS  
University of Southern  
California

ROBERT C. TRYON  
University of California

MICHAEL J. ZIGLER  
Wellesley College

JOSEPH ZUBIN  
Psychiatric Institute,  
Columbia University

If this space should be unstamped, this is the regular library edition.  
But if this space is stamped with a designating title, this is a special  
edition, sold under the restrictions of a bilateral contract, and may not  
be resold for a period of five years from date of publication.

## VOLUME 60

1965

Copyright, 1965, by The Journal Press  
Provincetown, Massachusetts, U. S. A.

All rights reserved

Published Quarterly at 35 New St., Worcester, Mass.

# AUTHOR INDEX

Adrian, Robert J. ....	81	Lerea, Louis .....	265
Allison, Roger B., Jr. ....	67	Lubetkin, Barry S. ....	313
Benton, Joseph G. ....	235	Luria, Zella .....	87
Bernard, J. L. ....	283	Macbride, Philip D. ....	9
Berry, Newell H. ....	31	McDonald, Robert L. ....	129
Bieliauskas, Vytautas J. ....	255	McGinnies, Elliott .....	1
Bronzo, Anthony F., Jr. ....	301	McLaughlin, Dennis .....	147
Brown, D. W. ....	295	McTee, A. C. ....	277
Buckhout, Robert .....	123	Meek, Edwin E. ....	165
Cautela, Joseph R. ....	117, 135	Nelson, Paul D. ....	31
Ceraso, John .....	239	Nelson, Willard H. ....	51
Clance, Pauline R. ....	271	Park, John N. ....	245
Cohen, David B. ....	51	Perry, Sally L. ....	55
Coyle, F. A., Jr. ....	283	Richter, Maurice N., Jr. ....	291
DiPalma, Giuseppe .....	155	Rogers, Miles S. ....	221
Dixon, T. R. ....	271	Sharp, Heber C. ....	205
Dufort, Robert H. ....	25	Shure, Gerald H. ....	221
Ferguson, Leonard W. ....	309	Smith, Anthony J. ....	39, 245
Fitzgerald, Maureen P. ....	181	Sprinthall, Richard C. ....	313
Friedman, H. D. ....	295	Stachnik, Thomas .....	199
Frost, B. J. ....	251	Stotsky, Bernard A. ....	55
Gentry, Harold W. ....	171	Taylor, Charles .....	193
Gough, Harrison G. ....	155	Tolchin, Gerald .....	239
Greenwood, Dennis I. ....	193	Tuddenham, Read D. ....	9
Helms, Donald L. ....	147	Vacchiano, Ralph B. ....	81
Hirschenfang, Samuel .....	235	Ward, Bruce .....	265
Hoar, Jere R. ....	165	Weitzner, Martin .....	71
Kenney, James B. ....	177	Wortz, E. C. ....	277
King, H. E. ....	103	Zagona, Salvatore V. ....	213
Kintz, B. L. ....	147	Zurcher, Louis A., Jr. ....	213
Korner, Ija N. ....	67		



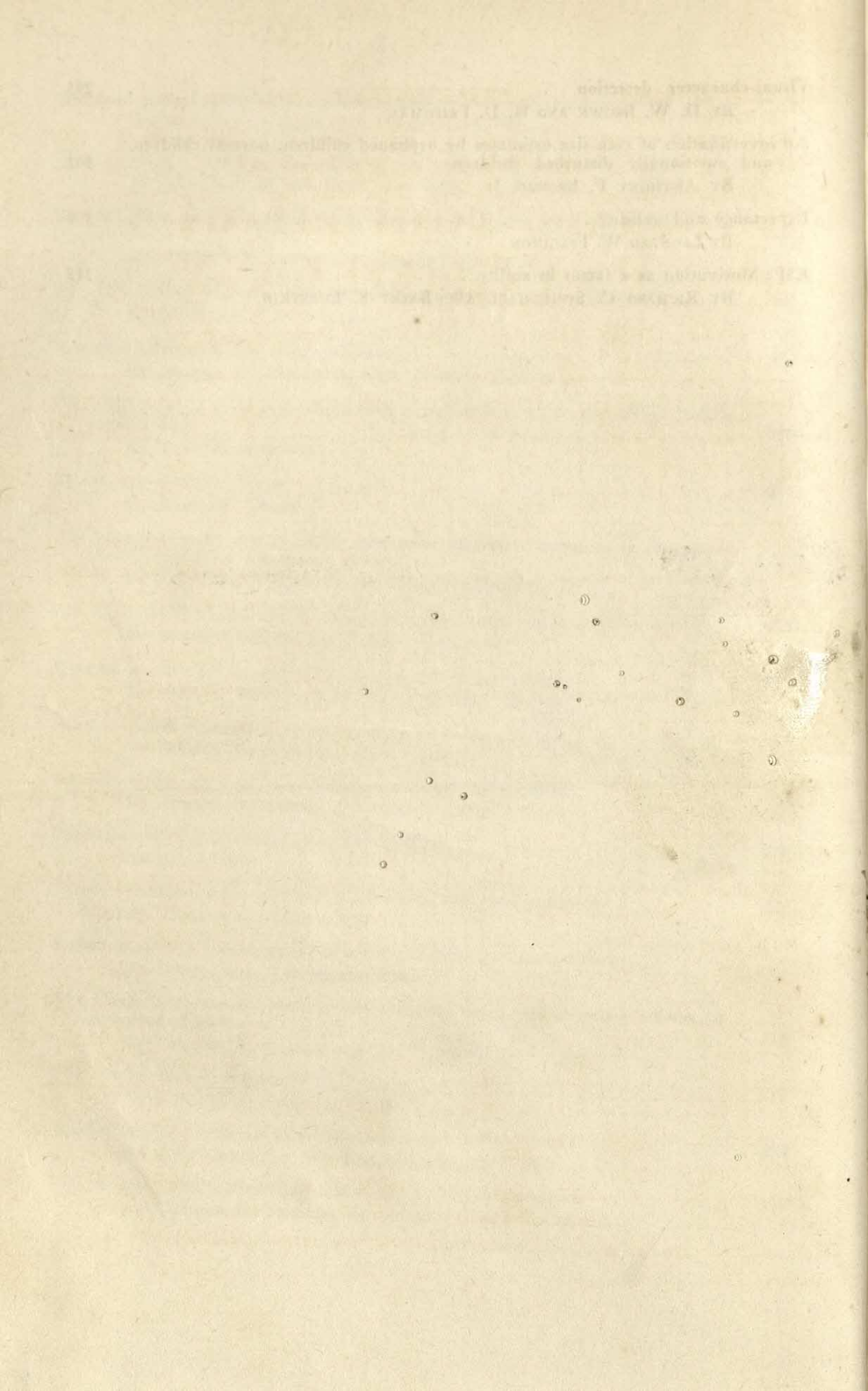
# TABLE OF CONTENTS

A cross-cultural comparison of printed communication versus spoken communication in persuasion . . . . .	1
BY ELLIOTT MCGINNIES	
The influence of self-confidence upon resistance of perceptual judgments to group pressure . . . . .	9
BY PHILIP D. MACBRIDE AND READ D. TUDDENHAM	
A suggested approach for the psychologist to the study of utopian writings . . . . .	25
BY ROBERT H. DUFORT	
The relationship between an individual's sociometric status in different groups over a two-year period . . . . .	31
BY PAUL D. NELSON AND NEWELL H. BERRY	
Developmental changes in cognitive balance . . . . .	39
BY ANTHONY J. SMITH	
Incidental stimuli: Some effects of color and magnitude in a conceptual situation . . . . .	51
BY DAVID B. COHEN AND WILLARD H. NELSON	
Type of reward, incentive, and incentive-sequence as factors in the motor performance of mentally retarded, physically handicapped, and college students . . . . .	55
BY SALLY L. PERRY AND BERNARD A. STOTSKY	
Comparative study of metaphoric thinking . . . . .	67
BY IJA N. KORNER AND ROGER B. ALLISON, JR.	
Manifest anxiety, amphetamine, and performance . . . . .	71
BY MARTIN WEITZNER	
The interrelation between the Picture Identification Test and the Activities Index . . . . .	81
BY RALPH B. VACCHIANO AND ROBERT J. ADRIAN	
Experiments on a postural autokinetic effect . . . . .	87
BY ZELLA LURIA	
The retention of sensory experience: IV. Short-delay versus long-delay intervals . . . . .	103
BY H. E. KING	
The influence of suggestion on the audioautokinetic effect . . . . .	117
BY JOSEPH R. CAUTELA AND DENNIS McLAUGHLIN	
Need for social approval and attitude change . . . . .	123
BY ROBERT BUCKHOUT	
Personality characteristics, cigarette smoking, and obstetric complications . . . . .	129
BY ROBERT L. McDONALD	
The problem of backward conditioning . . . . .	135
BY JOSEPH R. CAUTELA	
The effect of rate of presentation of paired associates on short-term retention . . . . .	147
BY DONALD L. HELMS AND B. L. KINTZ	

Attitudes toward colonialism, political dependence, and independence . . .	155
BY HARRISON G. GOUGH AND GIUSEPPE DI PALMA	
The semantic differential as a measure of subliminal message effects . . .	165
BY JERE R. HOAR AND EDWIN E. MEEK	
A comparison of the organizational climates of Negro and white elementary schools . . .	171
BY HAROLD W. GENTRY AND JAMES B. KENNEY	
The relationship between expressed self-esteem and assumed similarity . . .	181
BY MAUREEN P. FITZGERALD	
Adaptive testing in an older population . . .	193
BY DENNIS I. GREENWOOD AND CHARLES TAYLOR	
Contrast effects in extinction responding as a function of response effort during conditioning . . .	199
BY THOMAS STACHNIK	
Identifying placebo reactors . . .	205
BY HEBER C. SHARP	
The relationship of verbal ability and other cognitive variables to the open-closed cognitive dimension . . .	213
BY SALVATORE V. ZAGONA AND LOUIS A. ZURCHER, JR.	
An empirical evaluation of the effect of item overlap on factorial stability . . .	221
BY MILES S. ROGERS AND GERALD H. SHURE	
Delayed intellectual development in cerebral palsy children . . .	235
BY SAMUEL HIRSCHENFANG AND JOSEPH G. BENTON	
Selective trace arousal . . .	239
BY GERALD TOLCHIN AND JOHN CERASO	
Some correlates of inner directedness and other directedness . . .	245
BY ANTHONY J. SMITH AND JOHN N. PARK	
Subjective colors: An objective-color artifact . . .	251
BY B. J. FROST	
Recent advances in the psychology of masculinity and femininity . . .	255
BY VYTAUTAS J. BIELIAUSKAS	
Speech avoidance among children with oral-communication defects . . .	265
BY LOUIS LEREA AND BRUCE WARD	
The effect of previously learned verbal habits upon two response classes in verbal conditioning . . .	271
BY PAULINE R. CLANCE AND T. R. DEXON	
S-R relationships and transfer of training . . .	277
BY E. C. WORTZ AND A. C. MCTEE	
Logical thinking and paranoid schizophrenia . . .	283
BY F. A. COYLE, JR. AND J. L. BERNARD	
The concept of cognitive dissonance . . .	291
BY MAURICE N. RICHTER, JR.	



Visual-character detection . . . . .	295
By D. W. BROWN AND H. D. FRIEDMAN	
An investigation of coin-size estimates by orphaned children, normal children, and emotionally disturbed children . . . . .	301
By ANTHONY F. BRONZO, JR.	
Expectancy and validity . . . . .	309
By LEONARD W. FERGUSON	
ESP: Motivation as a factor in ability . . . . .	313
By RICHARD C. SPRINTHALL AND BARRY S. LUBETKIN	





\$20.00 per annum  
\$15.00 per volume  
\$7.50 per half volume

QUARTERLY  
Two volumes per year  
Immediate publication

May, 1965  
Volume 60, First Half

(Founded in 1935 by Carl Murchison)

# The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a liberal contract, and may not be resold for a period of five years from the date of publication.

MAY, 1965

(Manuscripts are printed in the order of final acceptance)

- A cross-cultural comparison of printed communication versus spoken communication in persuasion . . . . . 1  
BY ELLIOTT MCGINNIES
- The influence of self-confidence upon resistance of perceptual judgments to group pressure . . . . . 9  
BY PHILIP D. MACBRIDE AND READ D. TUDDENHAM
- A suggested approach for the psychologist to the study of utopian writings . . . . . 25  
BY ROBERT H. DUFORT
- The relationship between an individual's sociometric status in different groups over a two-year period . . . . . 31  
BY PAUL D. NELSON AND NEWELL H. BERRY
- Developmental changes in cognitive balance . . . . . 39  
BY ANTHONY J. SMITH

(OVER)

Copyright © 1965, by The Journal Press  
Provincetown, Massachusetts, U. S. A.

All rights reserved

Published at 35 New St., Worcester, Mass.  
Second-class postage paid at Worcester, Mass.

Send all mail to 2 Commercial Street, Provincetown, Massachusetts

Incidental stimuli: Some effects of color and magnitude in a conceptual situation . . . . .	51
By DAVID B. COHEN AND WILLARD H. NELSON	
Type of reward, incentive, and incentive-sequence as factors in the motor performance of mentally retarded, physically handicapped, and college students . . . . .	55
By SALLY L. PERRY AND BERNARD A. STOTSKY	
Comparative study of metaphoric thinking . . . . .	67
By IJA N. KORNER AND ROGER B. ALLISON, JR.	
Manifest anxiety, amphetamine, and performance . . . . .	71
By MARTIN WEITZNER	
The interrelation between the Picture Identification Test and the Activities Index . . . . .	81
By RALPH B. VACCHIANO AND ROBERT J. ADRIAN	
Experiments on a postural autokinetic effect . . . . .	87
By ZELLA LURIA	
The retention of sensory experience: IV. Short-delay versus long-delay intervals . . . . .	103
By H. E. KING	
The influence of suggestion on the audioautokinetic effect . . . . .	117
By JOSEPH R. CAUTELA AND DENNIS McLAUGHLIN	
Need for social approval and attitude change . . . . .	123
By ROBERT BUCKHOUT	
Personality characteristics, cigarette smoking, and obstetric complications <sup>1)</sup> . . . . .	129
By ROBERT L. McDONALD	
The problem of backward conditioning . . . . .	135
By JOSEPH R. CAUTELA	



## A CROSS-CULTURAL COMPARISON OF PRINTED COMMUNICATION VERSUS SPOKEN COMMUNICATION IN PERSUASION\*<sup>1</sup>

*Department of Psychology, University of Maryland*

---

ELLIOTT MCGINNIES

---

### A. INTRODUCTION

Many of the interesting problems in social psychology have antecedents or analogies in the experimental laboratory. Generally, the isolation of variables that control a specific bit of social behavior is accomplished more readily under laboratory than under other conditions. Not infrequently, however, the results of laboratory experiments are as inconsistent as are those obtained through experiments under naturalistic and less well-controlled circumstances. In such instances, theory construction seems premature until reliable sets of empirical data have been accumulated. The present experiment was designed to produce evidence bearing upon the problem of the relative effectiveness in persuasion of materials presented visually or auditorily. Prior research on this question, both in, and out of the laboratory, has yielded equivocal results.

A series of controlled experiments dealing with the effectiveness of radio and other media as instruments of vocal communication was reported some years ago by Cantril and Allport (1). Subsequent reviews of research in this area have been made by Hovland (4) and Klapper (6). A superiority of vocal over printed material for purposes of persuasion in general was found by Wilke (9) and Carver (2). Day and Beach (3) concluded that the advantage of the auditory mode was most apparent with meaningful and familiar material, whereas meaningless and unfamiliar material could be presented more efficiently visually. Silent reading, on the other hand, was reported by Knower (7, 8) to have an advantage over material read aloud by one member of a group in modifying attitudes.

\* Received in the Editorial Office, Provincetown, Massachusetts, on March 23, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was done under Contract Nonr 3720(01), NR 171-250 with the Group Psychology Branch, U.S. Office of Naval Research. A special debt of gratitude is acknowledged to Professor Katsuo Sano of Keio University, who supervised the collection of data from subjects at that university. Mr. Shunsuke Kamei handled the translations, and Miss Reiko Atsumi took responsibility for preparing the tape recordings and the printed materials.

The methodological problems of comparing different communication media have been pointed out by Hovland (4) who, nevertheless, concluded that the oral presentation of material seems to be more effective than visual presentation in changing opinions. Personal influence, of a face-to-face variety, has been reported by Katz and Lazarsfeld (5) to be more effective than any of the mass media in influencing opinions. One may conjecture that orally presented communications are more characteristic of personal influence and, as Cantril and Allport have suggested, permit greater opportunity for social facilitation to occur. Printed material, on the other hand, represents a more impersonal stimulus.

In assessing the research in this area, Klapper (6) concludes that "no mass medium can be assumed to be generally or always more persuasive than any other mass medium." He adds, "Some topics, furthermore, may be susceptible of better presentation by visual than oral means, or by print rather than by film, while for other topics no such difference exists."

This problem has been approached from another angle: namely, that of verbal learning and retention. Williams and Derks (10) have reviewed the status of research in visual modes versus auditory modes of presenting information and have reported the results of an experiment designed to determine the relative effectiveness of those two methods of presentation on the acquisition of paired associates. Williams and Derks' subjects learned three lists of paired-associates trigrams having different levels of pronounceability and association values under conditions of aural, visual, or combined presentation. As anticipated, increased pronounceability and association value resulted in an increased number correct. Both the visual and the combined presentations were more efficient than the aural. The superiority of the visual mode over the aural mode was greatest for highly pronounceable material having intermediate association values.

One is led to the conclusion that most of the studies that have grappled directly with the persuasion process favor auditory over visual communication. Williams and Derks' findings, however, imply that simple paired-associates learning is more efficient under conditions of visual than of auditory presentation. It is probable, of course, that connected prose of the sort necessarily used in persuasion is affected differently by mode of presentation than are trigrams. The experiment to be reported is an attempt to gain further insight into this matter in a seminaturalistic setting, by measuring the reactions of Japanese university students to persuasive communications on a topic with which they were familiar. The communications were presented to subjects by



tape-recording or in printed form. Both attitude change and evaluations of the communication and the communicator were measured under the two conditions.

## B. METHOD

### 1. *Subjects*

The subjects who completed the experiment were 55 male undergraduates and 13 female undergraduates enrolled in two psychology classes at Keio University in Tokyo. A number of subjects were lost due to their absence on the day of the experiment, and others failed to complete properly the questionnaire materials. There was no effective way to prevent the foregoing attrition, which appeared to be distributed randomly between the experimental groups. Difficulty in obtaining subjects for this experiment prevented use of a control group that would have been pretested and post-tested without exposure to a communication. However, experience with such controls in other experiments had shown us that no attitude change could be expected in control groups.

### 2. *Materials*

The issue used as a vehicle for persuasion was the official position taken by the United States after its discovery of missile sites and a weapons buildup in Cuba in the fall of 1962. A Likert-type scale was devised to measure the extent to which the respondent supported the actions taken by the United States: i.e., a naval blockade and insistence that the missiles be removed. Item analysis yielded eight statements that correlated satisfactorily with total score. Two items reflecting a neutral position were included, but were not scored. Each item provided seven response alternatives, ranging from "strongly agree" to "strongly disagree," so that total score for an individual could range from 8 (strongly anti-U.S.) to 56 (strongly pro-U.S.).

The persuasive communication, reproduced on magnetic tape and in printed form, was adapted from a speech by Ambassador Adlai Stevenson before the United Nations. The excerpts selected included a quotation from President Kennedy on the facts of missile-site development in Cuba and references to the responsibility of the Soviet Union in providing logistic support for the project. The remarks were translated into Japanese, checked for accuracy, and then tape recorded by a male student of dramatic arts at Nihon University in Tokyo. The recorded communication lasted approximately seven minutes. A written version in Japanese characters was also prepared and printed on two and one-half pages of 11"  $\times$  8½" paper.

### 3. Procedure

Arbitrarily, students in one of the classes were designated as recipients of the spoken communication and, at the conclusion of the experiment, consisted of 31 males and 11 females. The students in the other class, 24 males and two females, received the printed communication. All of the subjects took the attitude scale one week prior to the time that the experiment was to be conducted. To insure the anonymity of the subjects (and to increase the validity of replies), the questionnaires were identified only by number. A small slip of paper containing the number on the form was detached by the student and retained for later use.

On the day of the experiment, the subjects were told that the investigators wished to make a psychological study of reactions to a communication and to a communicator. Subjects were given a printed form containing a seven-point rating scale for judging the convincingness of a communication and nine pairs of bipolar adjectives describing a communicator. Each group of subjects was then exposed to the pro-U.S. spoken or printed communication on Cuba, but the source of the communication was not identified. After the subjects had rated the convincingness of the argument on the scale (which was anchored at the extremes by "very convincing" and "very unconvincing") and had checked the adjective in each pair that better described the speaker, they were asked to respond once more to the attitude scale. (The subjects identified their materials with the numbers assigned earlier.)

## C. RESULTS

### 1. Initial Attitude

Because the subjects had been assigned arbitrarily to the experimental conditions, the first step was to determine whether or not the two groups were similar in initial attitudes. The "listening" group had a mean attitude score of 39.67, not significantly different from that of 38.62 for the "reading" group. Since the point of exact "neutrality" on the scale is 32, the foregoing values indicate that both groups were initially "somewhat favorable" toward the stand taken by the United States in the Cuban crisis.

### 2. Attitude Change

Because the pre-experimental and postexperimental attitude scores are correlated measures, difference scores for each subject were computed. To determine whether or not sex differences were present in the "listening" group, which contained 11 females, the mean attitude change scores for males and



females were examined separately. The mean attitude change score for males, 1.00, does not differ significantly from that for females, 2.27; therefore, male and female scores for both groups are combined in subsequent analyses. The mean attitude-change score for the subjects in each of the experimental groups was tested by means of Student's *t* against a null hypothesis of zero change. The results are shown in Table 1.

TABLE 1  
ATTITUDE-CHANGE MEASURES

Medium	<i>N</i>	Mean	<i>SD</i>	<i>t</i>	<i>p</i>
Vocal	42	1.33	5.30	1.63	> .10
Printed	26	2.35	5.19	2.31	< .05

Subjects who *read* the persuasive communication moved a significant extent in the direction advocated. Subjects who listened to the tape-recorded argument, however, were not significantly influenced.

### 3. Convincingness Ratings

Each subject, immediately after exposure to the communication (printed or vocal) rated its convincingness on a seven-point scale. Mean convincingness ratings were computed for both the reading and the listening groups, and the two means were compared using the *t* test for uncorrelated measures with unequal numbers of cases. The results are shown in Table 2.

TABLE 2  
MEAN CONVINCINGNESS RATINGS

Medium	<i>N</i>	Mean	Difference	<i>S</i> <sub>Diff.</sub>	<i>t</i>	<i>P</i>
Vocal	42	4.36				
Printed	26	5.05	.68	.312	2.18	< .05

It is clear that subjects who *listened* to the communication tended to rate it in the middle of the scale at the position labeled "neither convincing nor unconvincing." Subjects who *read* the communication judged it "somewhat convincing," a rating significantly more positive than that accorded by listening subjects.

### 4. Perception of the Communicator

The nine pairs of adjectives from which the subjects were asked to choose those most descriptive of the speaker are susceptible to several forms of analysis. One objective in each pair could be designated "favorable"; the other,



"unfavorable." For example: honest—dishonest, interesting—uninteresting, intelligent—unintelligent. Because some of the subjects failed to select one adjective from each pair, the most appropriate measure seems to be the ratio obtained by dividing the number of favorable adjectives by the total number of adjectives checked. This ratio can vary from zero (indicating complete disapproval of the speaker) to 100 (indicating complete approval). The mean favorability ratios were compared to determine whether or not the medium of presentation of the communications is differentially related to evaluation of the speaker. The results of this analysis are shown in Table 3.

TABLE 3  
ADJECTIVE FAVORABILITY RATIOS

Medium	N	Mean F/T Index*	Differ- ence	$S_{Diff.}$	$t$	$p$
Vocal	42	65.60	7.86	2.54	2.54	< .02
Printed	24	73.46				

\* For explanation of index, see text p. 6.

Subjects who *read* the communication assigned a significantly greater proportion of favorable adjectives to the speaker than did those who *listened*. The fact that the communicator was perceived favorably by both groups probably reflects an initially favorable disposition toward the point of view advocated.

#### D. DISCUSSION

The results show a consistent advantage of the written communication over the spoken communication with respect to the three measured variables of attitude change, convincingness ratings of the arguments, and perceptions of the communicator; nevertheless, certain features of this type of situation pose difficulties of interpretation. For one thing, it is difficult if not impossible to control precisely the amount of exposure to the two modes of presentation. Undoubtedly, speed of comprehension of both types of material varies from one subject to another. In the present study, the experimenter allowed sufficient time for a single reading by the subjects who received the printed communication. This time was slightly longer than that allowed for presentation of the material by tape recording. Systematic control of exposure time could be achieved, but would require presentation of the communication on an individual rather than on a group basis.

The results run counter to most of those in prior reports on media effectiveness which, as noted earlier, suggest that live speakers, recorded speakers,

and print are effective in roughly that order. Several experiments conducted during the 1930s indicated a superiority of printed over oral presentation in producing retention of complex material, and this was also Williams and Derks' finding with simple materials (10). The present data suggest that visual presentations of persuasive communications are more effective in influencing Japanese students than are vocal presentations.

One explanation of the results is that the Japanese language is primarily a visual language. Written Japanese employs three types of characters: *katakana*, *hiragana*, and *kanji*. "Katakana" consists of phonetic symbols used in writing foreign names and words of foreign derivation, while "hiragana" is a syllabary used in all other cases. "Kanji" is an adaptation of Chinese ideographic script, and a list of 1850 characters was approved by the Diet in 1947 for use in newspapers, magazines, and official documents. It is possible that a greater wealth of meaning is conveyed by the printed than by the spoken language which, like English, has its share of ambiguities and vagaries. Because attitude change is essentially a learning process, such learning as occurs may be impeded or facilitated by those structural features of a language that distinguish its oral and written forms. This hypothesis is susceptible to experimental investigation.

Why was the printed argument rated as more convincing than the oral argument, and why did it reflect more favorably upon the communicator? One possible explanation relates to what several Japanese professors described as a resistance by Japanese students to external political pressures. Being allowed to read a persuasive communication may constitute a more acceptable situation because of the greater control this gives one over the stimulus material. When reading, one literally holds the communication in his hands and is not merely a passive recipient of a voice beamed over a loudspeaker. In a sense, the reader participates in the communication experience to a greater extent than the listener and thus may have a more positive "set" toward both the communication and the communicator.

The greater credibility of written material, the richness of meaning conveyed by printed as opposed to spoken Japanese, and the more active role played by a reader (as contrasted with that of a listener) are possible explanations of the results. The relative weights of these several alternatives remain to be determined.

#### REFERENCES

1. CANTRIL, H., & ALLPORT, G. W. *The Psychology of Radio*. New York: Harper, 1935.



2. CARVER, M. E. Listening versus reading. In H. Cantril and G. W. Allport. *The Psychology of Radio*. New York: Harper, 1935.
3. DAY, W. F., & BEACH, B. R. A survey of the research literature comparing the visual and auditory presentation of information. Technical Report No. 5921, United States Air Force, Air Material Command, Wright-Patterson Air Force Base, November, 1950.
4. HOVLAND, C. I. The effects of mass media of communication. In G. Lindzey (Ed.), *Handbook of Social Psychology*. Cambridge, Mass.: Addison-Wesley, 1954.
5. KATZ, E., & LAZARSFELD, R. F. Personal Influence, the Part Played by People in the Flow of Mass Communications. Glencoe, Ill.: The Free Press, 1961.
6. KLAPPER, J. T. The Effects of Mass Communication. Glencoe, Ill.: The Free Press, 1961.
7. KNOWER, F. H. Experimental studies of changes in attitude. I. A study of the effect of oral argument on changes of attitude. *J. Soc. Psychol.*, 1935, **6**, 315-347.
8. ———. Experimental studies of changes in attitude. II. A study of the effect of printed argument on changes in attitude. *J. Abn. & Soc. Psychol.*, 1936, **30**, 522-532.
9. WILKE, W. H. An experimental comparison of the speech, the radio, and the printed page as propaganda devices. *Arch. Psychol.* (N.Y.), 1934, No. 169.
10. WILLIAMS, J. M., & DERKS, P. L. Mode of presentation and the acquisition of paired associates that differ in pronunciability and association value. *J. Verb. Learn. & Verb. Behav.*, 1963, **2**, 453-456.

Department of Psychology  
University of Maryland  
College Park, Maryland



## THE INFLUENCE OF SELF-CONFIDENCE UPON RESISTANCE OF PERCEPTUAL JUDGMENTS TO GROUP PRESSURE\*<sup>1</sup>

*Department of Psychology, University of California, Berkeley*

PHILIP D. MACBRIDE AND READ D. TUDDENHAM

### A. PROBLEM

Visual perception is man's primary link to external reality. His survival depends upon its accuracy no less today than in the distant past. The velocity of an onrushing car on a busy highway offers as critical a test of perceptual judgment as the speed of a charging beast—and the distance to the nearest tree. Small wonder that one is spurred to action when he discovers a discrepancy between immediate experience and other evidence as to what is "out there."

Two principal modes of adjustment are employed in an effort to dispel uncertainty in visual experience. One takes a closer look,—and one turns to others, convinced that if they agree on what they see, he will have a valuable check upon his own perceptions. Axiomatic to one's understanding of self and others is the belief that these two prime sources of information concerning reality will confirm each other in all important respects. If one is alone in his view, contradicted by other observers, the uneasy thought may arise that the defect lies in oneself.

It is the authors' premise that such a contradiction results in a reduction of confidence in one's ability to perform the perceptual task in question. Such loss in confidence will not be equal in different individuals. It may be expected to vary in severity depending upon the attitude of the person toward the situation, toward the others, and toward himself. However, some reduction of confidence is hypothesized in every instance of unanimous opposition from others. Yielding to the norm of the others may be expected to follow this weakening of confidence in the self.

In the previous papers of this series (6, 7, 8, 9, 10, 11, 12, 13) various determinants of yielding were investigated. The present research concerned

\* Received in the Editorial Office, Provincetown, Massachusetts, on March 23, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This paper is an abridgment of Technical Report No. 10, ONR Project NR 170-159. The financial support of the Office of Naval Research is gratefully acknowledged; however the Navy Department is not responsible for the author's findings or conclusions.

the relationship between yielding and self-confidence. The authors hypothesized that yielding is an inverse function of confidence, and that confidence in one's perceptual judgments is in part dependent upon the accuracy with which one has made similar judgments in the past.

Two principal questions were investigated: (a) Do subjects who make many errors on a given perceptual task yield more to group pressure than do subjects who make few? (b) Do subjects who are forced to make errors on a training series and who are made aware of their inaccuracies subsequently yield more than do subjects who are aided to avoid errors in training trials?

To resolve the first question, the amount of yielding was related to degree of success at a perceptual task—in the present instance, a series of visual-tracing problems. The problems were hard enough that subjects differed appreciably in skill on a preliminary control series run without group pressure. It was predicted that persons who had made frequent errors on the control series would yield under group pressure more than subjects who had made few errors.

The second question had several aspects. Would compliant subjects respond independently following a success experience? Would independent subjects commence to yield following a failure experience? How much could individual differences in tendency to yield be modified by experimental means?

Varying the exposure time of the perceptual problems provided a simple and reliable means of controlling the subjects' success or failure. Two experimental conditions were employed. For one group, a lengthened exposure reduced errors to a minimum. Feedback to the subjects was provided in order to increase their self-confidence to make the perceptual judgments. This treatment was expected to reduce yielding when under pressure of a false group consensus. For a second group, a shortened exposure increased errors sharply. This treatment was expected to lead to lowered self-confidence and increased susceptibility to group pressure.

## B. METHOD

### 1. *Subjects*

Fifty-two men and 50 women undergraduates at the University of California served as subjects. Persons who reported poor eyesight or who had any prior knowledge of the investigation were excluded. All subjects were run in groups of five members of the same sex.

### 2. *Apparatus*

The apparatus has been described elsewhere (7). Five subjects sat next to each other, each at a separate panel screened from his neighbors, and responded



to visual problems projected on the wall from slides. Subjects were told that on practice trials they might answer in unison, but that on test trials they were to answer one after the other "to permit automatic recording of the answers." The prescribed order of response, changing from trial to trial, was transmitted to each panel by illuminating one of five red lights in a column at the left of the panel. A row of nine numbered toggle switches below the panel permitted each subject to make his responses. Five rows of nine green lights each, ostensibly intended to show him when to take his turn, signalled to each subject the switches supposedly chosen on the other panels for the trial in question. In *fact*, the experimenter controlled all the lights. He thus could lead each subject to believe that he was the last to respond and that the four others in his group had all chosen a response that appeared to him to be phenomenally incorrect.

### 3. *Items*

All stimulus items were visual mazes of the same general type and similar to one another in appearance. Each maze consisted of five wavy lines that crisscrossed each other in an irregular fashion. One line originated at an X symbol and terminated with one of the digits from 1 to 9. The other four lines linked random pairs of the remaining eight digits. The subject's task was to discover which digit was linked to the X and to indicate his choice by pressing the switch of the corresponding number when it was his turn to respond.

Mazes were projected from slides onto a screen about 12 feet in front of the subjects, so that the image occupied a 30-inch square. Mazes were shown white on a dark ground and produced a clear and lively picture.

Success at the task depended largely upon the length of the exposure, which was accurately regulated by an adjustable timer. Preliminary experimentation established that a three-second exposure produced a satisfactory range of individual differences in error score. A four-second exposure lowered the frequency of errors below 10 per cent; a two-and-one-half-second exposure raised errors to twice the three-second frequency. These differences in exposure time were small enough to go unnoticed by the subjects, a crucial consideration in this experiment.

### 4. *Procedure*

The purpose of the experimental series of trials was to provide a measure of each subject's skill at the perceptual task and of his susceptibility to group pressure in the form of judgments apparently diverging from his own. A second purpose was to provide a measure of *change* in his susceptibility, in consequence of an intervening reality-testing experience designed to enhance self-



confidence in half the subjects and to weaken it in the other half. The series consisted of 49 trials divided into six parts as follows:

1. Instructional series of four slides to familiarize subjects with the task and the apparatus.

2. First control series (C1) of 10 slides, the last five constituting the critical trials. Subsequently, this same set of five slides was repeated in a reality-testing condition, in two test series to gauge susceptibility to pressure, and in a second control series. The need for accuracy was strongly emphasized, but subjects were told, "Since this is a practice series the recorder will not be operating and you need not answer in turn." Slides were exposed for three seconds, with feedback to each panel only of that subject's own response. Following the series, each person filled out a card to indicate how many errors he estimated he had made on the 10 trials.

3. First test series (T1) of 11 slides. Subjects were told to answer in turn "to permit recording." Slides were exposed for three seconds. On the critical trials (namely, 3, 5, 6, 8, and 10) each person was assigned to the fifth turn, and a false but ostensibly unanimous choice by "the other four subjects" was simulated by the experimenter. On trials 1, 2, 4, 7, 9, and 11 the correct choice was supplied.

4. Reality-testing series of eight slides, including repetition of the five critical slides from the preceding series. These trials were described as "additional practice." Subjects answered with no knowledge of other subjects' responses. Half of the groups were given a two-and-one-half-second exposure to each slide (the R— treatment), too brief for success much above chance levels. The remaining groups were given a four-second exposure to each slide (the R+ treatment) permitting 90 per cent accuracy. Immediately after each trial the slide was shown again for 10 seconds, so that each subject could verify or correct his choice. This procedure demonstrated to each R— subject his own fallibility and to each R+ subject his own competence. As after the control series, each subject estimated how many errors he had made and recorded his estimate on a card.

5. Second test series (T2), the slides and experimental conditions identical with those of the first test series.

6. Second control series (C2), consisting of the five critical slides presented under the same conditions as in the first control series (C1). A set of questionnaires [see (12)] completed the session.

The initial control series provided a measure of unpracticed ability to cope with the maze problem, a measure that was correlated with yielding score to determine the degree to which yielding is a function of accuracy. Improvement from practice or decrement from fatigue was evaluated by comparing scores of the first and second control series. The average of the two control runs

indicated the accuracy of the subjects under nonpressure conditions with practice and fatigue effects balanced out, and provided the most accurate value for comparison with yielding score. The mean number of errors on the series of test trials for which erroneous false norms were supplied was compared with the mean number of errors on the control trials to measure yielding. Differences in susceptibility to the false norm between the first and the second test series were attributable to the increase or decrease in self-confidence engendered by the reality-testing series that intervened.

### C. RESULTS

#### 1. *Did Yielding Occur?*

Table 1 compares distributions of perceptual errors on Control Series 1 and 2 with distributions of yielding errors (i.e., errors in agreement with the false

TABLE 1  
FREQUENCY DISTRIBUTIONS FOR FIVE CRITICAL TRIALS OF PERCEPTUAL ERRORS ON  
CONTROL SERIES AND YIELDING ERRORS ON TEST SERIES

Number of errors	Men (N = 52)				Women (N = 50)			
	Control series (C1)	Test series (T1)	Test series (T2)	Control series (C2)	Control series (C1)	Test series (T1)	Test series (T2)	Control series (C2)
5		4	4			3	6	
4		2	3		1	3	6	
3*	1	5	9		1	9	4	
2	4	12	3	1	5	5	8	
1	16	15	13	11	19	14	7	12
0	31	14	20	40	24	16	19	38
Mean	0.5	1.6	1.5	0.2	0.7	1.6	1.8	0.2
Per cent	10	32	30	5	14	32	36	5

group norm) on Test Series 1 and 2. All distributions are calculated for the five critical slides only and therefore are directly comparable. Several facts are immediately evident.

Yielding errors occurred much oftener than did perceptual errors on the same problems presented with the same exposure time. Subjects faced with a unanimous though incorrect consensus yielded roughly one-third of the time. In contrast, perceptual errors made under nonpressure conditions varied from 14 per cent in the initial control series to five per cent after practice. "Independent" errors (i.e., errors on critical trials that differed from the false norm) were rare—two per cent of the men's responses, four per cent of the women's.

A difference score of yielding errors minus perceptual errors could be com-



puted since each subject responded to the same slides under pressure ( $T1 + T2$ ) and without pressure ( $C1 + C2$ ). The mean difference score for men was 2.3; for women, 2.4. The  $t$  values were 5.5 and 5.8, respectively, both beyond the one per cent level of confidence. Clearly, considerable yielding occurred.

Individual differences in tendency to yield were large, extending over the entire range of possible scores. This finding is in agreement with that of earlier work (1, 4, 7), as is the finding that scores tended to pile up at the non-yielding end of the continuum; but whereas 79 per cent of the men and 74 per cent of the women made no more than one error in 10 control trials ( $C1 + C2$ ), only 38 per cent of the men and 44 per cent of the women made no more than one error on the 10 pressure trials ( $T1 + T2$ ).

The results for men and for women are remarkably similar, an unexpected finding in view of the sizeable sex difference among college students reported by Crutchfield (4) and Tuddenham (7, 13); however there was a sex difference in the correlation between yielding and other variables, as described later in this report.

Errors on the second control series were fewer than half as frequent as on the first control series. This effect of practice in coping with the perceptual problem under nonpressure conditions is significant. (The  $t$  for men was 2.4; for women, 4.0). In contrast, the total number of errors on the second test series ( $R+$  and  $R-$  groups combined) was as great as on the first test series. This finding shows that group pressure led to yielding responses in spite of veridical perceptions.

## 2. *Was Yielding a Function of Confidence in Ability to Cope with the Task?*

• Both men and women showed marked individual differences in accuracy on the first control series, error scores ranging from zero to eight for 10 trials. Subjects' estimates of the number of errors made varied over about the same range. The correlation between errors actually made and error estimates was quite high ( $r = .81$  for men,  $.61$  for women), indicating that subjects had a fairly accurate idea of their own ability to cope with the problem. The authors had expected these error scores and error estimates under control conditions to be highly correlated with yielding scores. The findings were not consistent. For women, yielding correlated .32 with control-series error scores and .46 with error estimates. For men, the values were .02 and .00 respectively.



### 3. *Did Reality Testing Modify the Tendency to Yield?*

The primary objective of this experiment was to modify subjects' initial susceptibility to group pressure, by providing each one with an opportunity to determine his own ability at the task before he faced the pressure of the erroneous majority for a second time; however the reality-testing situation was manipulated so as to cause some subjects to commit a large number of errors and to cause others to experience error-free or nearly error-free performance. Although all members of a group of five necessarily received the same reality-testing experience regardless of individual accuracy or independence of judgment, there is no doubt that the change in exposure time effectively changed error scores on the reality-testing series. The frequency of error was only 12 per cent for men and 10 per cent for women given the four-second exposure time for the eight slides (R+ treatment). The frequency of errors was 50 per cent for men and 58 per cent for women given the two-and-one-half-second exposure time (R- treatment). The level of 50 per cent error was high enough to demonstrate to the subjects their fallibility. The 10 per cent error level under the R+ treatment was higher than that anticipated from preliminary work. Nevertheless about half the men and half the women in R+ groups had zero error scores, and only a handful made more than two errors.

Subjects' estimates of error were congruent with their actual scores. When given the opportunity to check the solution during the re-exposure of each slide, subjects were very accurate in estimating how many errors they had made. The great majority of subjects in long-exposure groups believed they had made fewer errors on the reality-testing series than on the first control series. With few exceptions, subjects in short-exposure groups believed they had made more. For long-exposure and short-exposure subjects combined, the correlation between error estimates and error scores approached unity.

What, then, was the effect of the reality-testing experience on susceptibility to group pressure in the second test series?

For each subject, the error score on the test series prior to reality testing (T1) was subtracted from the error score on the test series following reality-testing (T2). Table 2 demonstrates that there was a significant difference between long-exposure and short-exposure groups in mean change score. The  $t$  for men was 2.4; for women, 2.2. Thus it is clear that the error scores imposed on the subjects by extending or curtailing the exposure times of the mazes did affect subsequent behavior, although assignment to R+ or R-

TABLE 2  
FREQUENCY DISTRIBUTIONS OF CHANGES IN SCORE UNDER GROUP PRESSURE, FOR  
GROUPS GIVEN INTERVENING POSITIVE OR NEGATIVE REALITY TESTING

Change score <sup>a</sup>	Men		Women	
	Long- exposure group (R+) (N = 25)	Short- exposure group (R-) (N = 27)	Long- exposure group (R+) (N = 26)	Short- exposure group (R-) (N = 24)
5				
4				1
3		2	1	1
2	2	1	0	4
1	4	7	7	4
0	6	11	7	10
-1	7	6	9	4
-2	5		1	
-3	0		1	
-4	1			
-5				
Mean	-0.5	0.3	-0.2	0.6
SD	1.4	1.1	1.2	1.3
$t_1^b$	1.8	1.6	0.8	2.2*
$t_2^b$		2.4*		2.2*

<sup>a</sup> Change score = errors on second test series (following reality testing) minus errors on first test series.

<sup>b</sup> $t_1$  = significance of difference of mean from zero.

<sup>c</sup> $t_2$  = significance of difference between means of long-*vs.* short-exposure groups.

\* Significant at the .05 level. All  $t$ 's are evaluated as one-tail tests inasmuch as direction of difference is predicted by hypothesis.

treatment was random and although the R+ and R- groups were essentially equal in initial ability to cope with the task.

Table 2 also presents  $t$  tests of the divergence of the mean change score of each group from zero. These tests assess the significance of a change in yielding errors in consequence of reality testing for positive-treatment and negative-treatment groups considered separately. All  $t$ 's are evaluated as one-tail tests, inasmuch as the direction of observed differences is predicted by hypothesis. A small sex difference emerged. For the men, the principal effect was a *decrease* in yielding following positive reality testing. For the women, however, the greater shift was an *increase* in yielding following negative reality testing. The net result of the two trends was to make the average yielding score for women greater than that for men on the second test series, though there was no difference on the first (*see* Table 1). This descriptive finding, though suggestive of important differences between men and women in the psycho-



logical basis of yielding to pressure, should be replicated before it is regarded as established fact.

#### 4. *How Important Was Reality Testing as a Determinant of Subsequent Tendency to Yield?*

Data in the top row of Table 3 express in correlational terms the findings given in Table 2: namely, that there was a significant relationship between

TABLE 3  
CORRELATIONS AMONG ERROR SCORES ON DIFFERENT SERIES

Variable	Error score on reality-testing series (R+ and R— combined)		Estimated errors on reality-testing series (R+ and R— combined)		Error score in First test series (T1)	
	Men	Women	Men	Women	Men	Women
Change score (T2 — T1)	.48**	.43**	.58**	.48**	—	—
Error score (T2)	.20	.27*	.20	.35**	.64**	.68**

\* Significant at the .05 level (one-tail test).

\*\* Significant at the .01 level (one-tail test).

Both the number of errors estimated and the number of errors actually made in reality testing on the one hand and change in susceptibility to group pressure on the other. The correlations were lower but still positive when the reality-testing results were compared not with change score but simply with error score in the group-pressure situation following reality testing.

The largest correlations in error scores were those between the two test series where group pressure was imposed. The values (.64 for men and .68 for women) are both significant beyond the .01 level. It seems clear that although the experimental treatment of reality testing did significantly modify susceptibility to group pressure, nevertheless the degree of susceptibility or independence that each subject brought to the experiment was the more important determinant of his final behavior. When one considers the brevity of the reality-testing situation in comparison with a lifetime of confidence-building or confidence-breaking experiences prior to the experiment, it is not surprising that enduring attitudes toward the self continued to be of major importance in spite of our attempt to modify them experimentally for this particular task.

#### 5. *Questionnaire Data*

The rating scales completed by each subject at the close of the experimental session were designed to provide data on three main topics: (a) subjects' motivations and impressions of the experimental conditions; (b) subjects' de-



scriptions of their performances on the control and the test series; and (c) subjects' descriptions of interactions with "others."

In general, the subjects' self-reports bore out the proposition that confidence in ability to perform the task may be inferred from error scores. Correlations between errors during the reality-testing series and ratings of confidence were high and negative for both men and women: that is, when the reality-testing experience demonstrated to subjects a large number of errors, they reported a reduction of confidence in accuracy. Negative correlations between errors during reality testing and agreement with the statement "Practice trials increased my confidence" substantiate this same point.

The results on the rating scales pertaining to self and others indicated a high degree of realism in self-perceptions of performance. Subjects who yielded were aware of that fact. Responses to the item "Influenced by the answers of the others" correlated .60 and .68 with actual yielding scores for men and women respectively. Yielders indicated that they "paid close attention to the answers of the others." The more independent subjects "found the answers of the others unreliable as a guide for own answers." And, though most subjects reported their answers to be "often different from the others," agreement with this statement was rated with more intensity by the more independent, consistent with their lower yielding scores. Finally, the high yielders reported less confidence in their accuracy. The correlations were  $-.56$  for the men and  $-.49$  for the women; thus the self-reports of the subjects corroborate the hypothesized relationship between confidence and yielding.

#### D. DISCUSSION

In the present research, as in the previous studies in this series, the disparity in judgment between self and others much increased nonveridical reporting compared with that found under nonpressure conditions. The surprising result, in the light of previous experiments, is not the fact that one in three responses showed yielding—a value that corresponds closely to that observed by Asch in his classic studies (1, 2)—but that yielding responses were not more numerous.

Two important features of the present study operated to produce more yielding than Asch's procedure. (a) The task, while extremely simple with prolonged inspection, presented difficulties under the limitation of three-second exposure and demanded close attention to avoid error. (b) The stimulus figure was no longer visible at the moment of report. Checking one's answer after noting the disparity between self and others was impossible. Deutsch and

Gerard (5) reported that such a memory variation increased yielding. Given these conditions, the fact that two out of three of the responses under pressure were made in opposition to the group consensus attests to a strong motive or attitude of independence.

There were, however, large individual differences in the ability of the subjects to resist the group. Clearly, a major role in the determination of these differences was played by characteristics or traits that were a more or less permanent part of the subject and not developed as a result of immediate experimental conditions. Some insights into the personality configurations that make for yielding or independence have been offered by Barron (3), Crutchfield (4), and Tuddenham (9, 12).

In spite of the manifest resistance to yielding and in spite of the obvious importance of long-developed personal attitudes as determinants of yielding, the subjects responded to the immediate situation and "learned" from the reality-testing experience. Manipulation of the experimental conditions modified the tendency to yield; however the effect was relatively small and left initial individual differences more or less as they had been, as shown by large correlations of .64 for men and .68 for women between yielding scores on the first *vs.* the second test series. Clearly, individual differences stemming from stable characteristics of the self were more influential than situational variables introduced by the experimenter.

The foregoing observation is complicated, however, by a sizeable and unexplained sex difference that emerged. For women there was a correlation of .46 between number of errors made under pressure and errors made in the initial control series. For men, the same correlation was zero. It seems that experience with the task played a significant role at the outset for women in determining who would yield and how much. On the other hand, the men appeared to be oblivious to their own errors in adjusting to the false consensus on the first test series.

There exists much evidence that yielding to others may not involve the self in the same manner for women as for men. In American culture, independence for men is closely integrated with manliness and masculine maturity. It is a sign of weakness to give in. Conversely, a certain degree of docility and orientation to others is regarded as a feminine virtue, and seems consistent with the women's greater responsiveness to their initial experience with the mazes. To some extent, women deferred to the answers of the "others" if they had made many errors on the previous practice series.

If submitting to the majority is less vital a matter and less disruptive to



the self for women than for men, one would expect women to yield more frequently than men do. Both Crutchfield and Tuddenham reported significantly greater yielding among college women than among college men; however, in the present study virtually no sex difference was found. The mean yielding scores were 31 per cent for men and 34 per cent for women. With an older age group, Tuddenham found no sex differences in yielding and concluded that "at present, sex differences in yielding are not altogether disentangled from other variables. Further work . . . may help to resolve the matter" (7, p. 237). This conclusion applies equally well to the present investigation.

When one turns from general trends to individual behavior, the sharp contrasts in performance attest further to the important role of personal traits and attributes in shaping the reaction of the individual to unanimous contradiction by others. The authors anticipated that a reality-testing situation that provided strong evidence in support of an individual's ability to cope with a problem might increase his self-confidence enough for him to withstand the disagreement of others, in spite of a previous predilection for submitting to their judgments. The reality-testing series gave each individual an opportunity to check his answers. The time for checking was long enough to provide all the information needed to verify the answer. The subjects could have gained little more if they had been allowed to walk up to the screen and trace the winding pathways with their fingers. Most subjects looked away long before the figure left the screen. That their reality testing was true and accurate is shown by the high correlations between actual errors and estimates of errors. The failure of these correlations to reach 1.0 is probably attributable to faulty memory—not to failure to verify judgments made during the reality-testing series. It seems safe to conclude that subjects were provided a clear and adequate demonstration of individual ability to cope with the task.

Would not a subject who had repeatedly confirmed his own judgments, in thorough re-examination made after brief viewing, be better able to resist the erroneous majority in a subsequent test of yielding involving the same kind of judgments? Would not the reality-testing experience be "therapeutic" in restoring confidence and replacing imagined personal defects or inadequacies with solid assurance of proved ability? The records of individual subjects presented in Table 4 show that no such simple predictions are possible, and that demonstrated accuracy of judgment is not necessarily an antidote to yielding.

Male subjects A and B showed on the first practice series their ability to render accurate judgments when not under pressure, but both yielded on every trial when their answers were "disputed" by the unanimous majority. Both

TABLE 4  
RECORDS OF INDIVIDUAL SUBJECTS ON CRITICAL TRIALS

Series	Number of errors					
	A	Men B	C	Women D	E	F
First control (C1)	0	0	0	0	1	1
First test (T1)	5	5	2	1	0	0
Long-exposure reality-testing (R+)	0	0	0	—	—	—
Estimated errors on R+	0	0	0	—	—	—
Short-exposure reality-testing (R—)	—	—	—	7	6	6
Estimated errors on R—	—	—	—	7	7	5
Second test (T2)	1	5	1	0	4	1
Second control (C2)	0	0	0	1	0	0
Change score (T2 — T1)	—4	0	—1	—1	4	1

subjects were members of a positive reality-testing group, and both made no errors on that series of eight trials. There the similarity ended.

Subject A changed markedly in tendency to yield following the reality-testing series. After yielding to the group on the first critical slide of the second test series, he ceased to yield and responded veridically for the remainder of the series.

Subject B, on the other hand, responded to the contradictory consensus after the reality-testing series just as he had before. If the reality-testing experience had any effect on him, it did not show up in his performance. Repeated confirmation of the correctness of his judgment did not enable Subject B to resist the erroneous group norm. With the pressure off on the second control series, however, Subject B's performance became veridical again.

Male subject C was a typical member of the positive reality-testing group. After a record of no errors on the reality testing series, Subject C's yielding score decreased; but it remained similar in magnitude to his score on the first test.

Women subjects provided equally striking examples of contradictory reactions. All three of the women subjects cited in Table 4 were in groups given the negative reality-testing treatment. All three made many errors on the eight slides of that series. This fact was demonstrated to the subjects by a prolonged viewing of each slide. That the subjects were fully cognizant of their errors is shown by their error estimates. Yielding by Subject E increased from zero to four (out of five) as a result of this negative reality-testing. Subject D showed a reverse effect. She was able to reject the consensus of the "others" despite having just "failed" on seven of eight practice problems.

Subject F was typical of those individuals who found themselves in the negative reality-testing group. She yielded for the first time following the



negative series but, as with most of the other subjects, her performance on the second test series was very similar to that on the first.

All six of the foregoing subjects demonstrate a characteristic of the entire group; namely, the ability to resume error-free or nearly error-free judgments on the final control series regardless of performance under pressure. The stresses and tensions generated by yielding to an erroneous majority or by steadfastly maintaining one's independence from it did not disturb subsequent veridical performance when the pressure was removed.

### E. SUMMARY

Fifty-two college men and 50 college women participated in an investigation of the relationship between confidence and resistance to group pressure of perceptual judgments on a series of visual line-tracing problems. An electrical communication device confronted each subject in a group of five with unanimous wrong answers, purportedly coming from the four other members of the group. Each subject was made to believe he was last in turn to answer. The experimenter simulated the responses ostensibly coming from the others.

Approximately 30 per cent of the judgments conformed to the false consensus. As in previous studies of yielding, individual differences were large, with the majority of subjects concentrated toward the independent pole of the susceptibility continuum.

A reality-testing series, which allowed subjects to check on the accuracy of their judgments on a number of nonpressure trials, was effective in modifying subsequent yielding scores. In this training series, which was interpolated between two test series, subjects' error scores were manipulated experimentally to produce low error scores (positive reality testing) or high ones (negative reality testing). A difference score that measured change in yielding from the first to the second test correlated significantly with error scores on the reality-testing series. A changed degree of confidence in ability to perform the perceptual task was found to be the variable determining the shift in yielding and was clearly evidenced in subjects' estimates of their own errors.

The authors conclude that a demonstration of ability to perform a task can prepare an individual better to withstand group pressure, and that a demonstration of inaccuracy can weaken self-confidence and reduce ability to withstand pressure. The two tests of yielding, however, before and after the interpolated series, had a high correlation with each other, showing that such demonstrations did not substantially obscure initial individual differences in tendency to yield. Undoubtedly, stable personality characteristics are the primary determinants of yielding behavior.

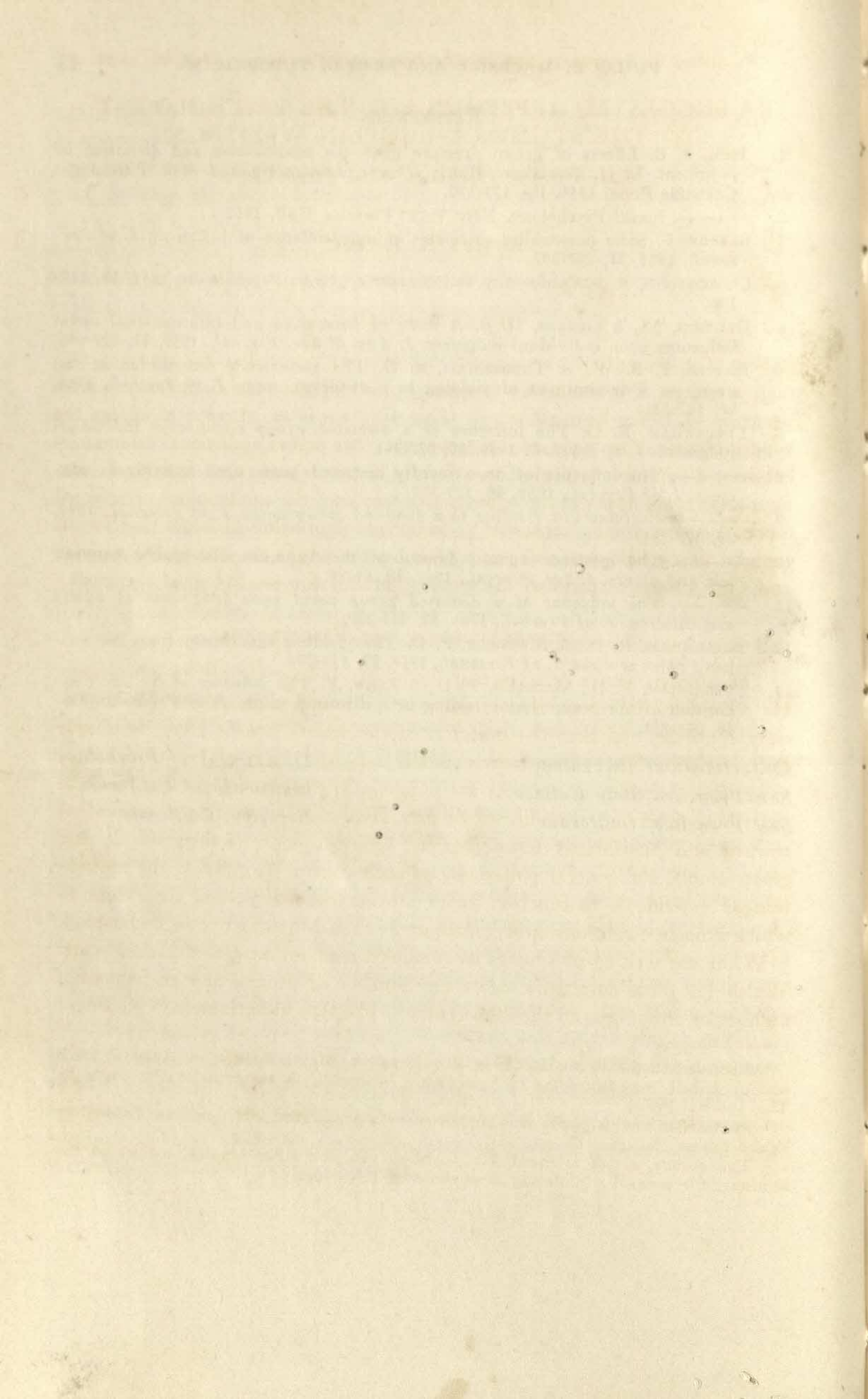
## REFERENCES

1. ASCH, S. E. Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, Leadership and Men*. Pittsburgh: Carnegie Press, 1951. Pp. 177-190.
2. ———. *Social Psychology*. New York: Prentice Hall, 1952.
3. BARRON, F. Some personality correlates of independence of judgment. *J. of Personal.*, 1953, **21**, 287-297.
4. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191-198.
5. DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. Abn. & Soc. Psychol.*, 1955, **51**, 629-636.
6. HARPER, F. B. W., & TUDDENHAM, R. D. The sociometric composition of the group as a determinant of yielding to a distorted norm. *J. of Psychol.*, 1964, **58**, 307-311.
7. TUDDENHAM, R. D. The influence of a distorted group norm upon individual judgment. *J. of Psychol.*, 1958, **46**, 227-241.
8. ———. The influence of an avowedly distorted norm upon individual judgment. *J. of Psychol.*, 1958, **46**, 329-338.
9. ———. Correlates of yielding to a distorted group norm. *J. of Personal.*, 1959, **27**, 272-284.
10. ———. The influence upon judgment of the apparent discrepancy between self and others. *J. Soc. Psychol.*, 1961, **53**, 69-79.
11. ———. The influence of a distorted group norm upon judgments of adults and children. *J. of Psychol.*, 1961, **52**, 231-239.
12. TUDDENHAM, R. D., & MACBRIDE, P. D. The yielding experiment from the subject's point of view. *J. of Personal.*, 1959, **27**, 259-271.
13. TUDDENHAM, R. D., MACBRIDE, P. D., & ZAHN, V. The influence of the sex composition of the group upon yielding to a distorted norm. *J. of Psychol.*, 1958, **46**, 243-251.

Department of Psychology  
San Francisco State College  
San Francisco, California

Department of Psychology  
University of California  
Berkeley, California





## A SUGGESTED APPROACH FOR THE PSYCHOLOGIST TO THE STUDY OF UTOPIAN WRITINGS\*<sup>1</sup>

*Department of Psychology, Wake Forest College*

---

ROBERT H. DUFORT

---

### A. INTRODUCTION

Recently, the author participated in a seminar devoted to the study of utopian writings. The seminar was given as part of the interdisciplinary honors program of Wake Forest College. This program is an attempt to expose the superior student to ideas and concepts that cut across traditional boundaries between disciplines; typically, two faculty members are asked to serve as senior participants in a weekly seminar that explores the implications of selected topics. Because utopias represent an important expression of issues that are of concern to psychologists, this paper describes the works covered by the seminar and some of the features of the work that call for psychological comment.<sup>2</sup>

### B. CONTRIBUTIONS OF THE PSYCHOLOGIST

The utopian writer believes that his society is best for all its inhabitants. He recognizes, therefore, that means must be found to guarantee that the society continues and that the members of the society benefit from its unique characteristics. Since these means usually take the form of behavior-control techniques (whether or not they are so labeled), the psychologist has much to contribute to an understanding of utopias. He can give attention to traditional reward and punishment techniques for behavior control, their use in any given utopia, the general philosophy of such control (e.g., 17), the motivation of individuals to function under utopian conditions, and the extent to which utopian conditions meet human needs as such needs are now understood.

While the writing of utopias has occupied man for several thousand years, Walsh (21) has detected a decreasing number of utopias and an increasing number of antiutopias in the 20th century. The typical antiutopia [e.g., *Brave*

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 6, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> Publication of this paper was supported by a grant from the Graduate Council of Wake Forest College.

<sup>2</sup> The author wishes to thank Dr. G. M. Bryan, the other faculty member in the seminar, for several stimulating discussions of this topic.



*New World* (8) and 1984 (15)] reflects pessimism about man's future, in contrast to the hopeful picture painted by the typical utopia. In particular, the utopia implies that man can use his enlightened nature to plan a better society, while antiutopias are expressions of doubt that man can rationally structure society for his betterment. Walsh attributes this change in the emphasis of imaginary societies to loss of hope for the peaceful and harmonious world that conceivably could be brought about by the proper use of material abundance, atomic energy, the United Nations, and medical advances against disease. In the antiutopia, the behavior-control techniques generally are more numerous and obvious than in the utopia. The basic assumption about human nature made by authors of both kinds of societies is that man is plastic and can be changed by experience to fit any environment. The common belief that the utopia uses reward and the antiutopia uses punishment to control behavior is an oversimplification, for one finds punishments in utopias [e.g., *The Republic* (16)] and rewards in antiutopias [e.g., *Brave New World* (8)]. It is not the kind of behavior-control device but the intention of the author that determines how a society should be regarded; the utopia is intended to be a society that is superior to any that has existed, and the antiutopia is intended to be a society inferior to any that has existed (21).

#### C. UTOPIAS

Perhaps the best introduction to utopias for the psychologist is Skinner's *Walden Two* (18). This well-known contemporary work is effective in stimulating interest and includes many of the issues that historically have concerned authors of imaginary societies (e.g., the question of exchanging some degree of personal freedom for happiness and the difference between persons born in the utopia and those joining it from the outside world). It contains such devices for behavior control as child-rearing practices that shape the individual to fit the society and censorship through the selection of reading material. Finally, *Walden Two* raises the fundamental issue of whether the behavioral norms of society should be determined empirically or by traditional value standards.

An appropriate second reading is Plato's *Republic* (16). Written around 390 B.C., *The Republic* is the first historically important utopia; it has influenced virtually all subsequent writings in this vein. It portrays an ideal state designed to provide justice for all inhabitants. Various means are proposed to establish and maintain this state; a selective education system, by which each person is trained to fit his ultimate place in the society, is the key process.

State regulation of private life and the subordination of the individual to the welfare of the whole are common. Treatment of the individual can be callous at times if the aims of the state are served. For the guardians, the ruling class that runs the state, a communistic form of existence is prescribed.

Utopian writing did not prosper between the time of the ancient Greeks and the Renaissance (13, 14, 21). A few works appeared that are partially in this tradition, such as St. Augustine's *City of God* (20); but the main line of utopian writing was in eclipse during this time. More's *Utopia* (12), published in 1516, marks the return of the major utopian works. It contains a curious combination of the good life and such stringent behavior controls as slavery, restricted travel, population control, and the death penalty for officers of government who discuss official matters outside the legislative body. In spite of these ominous features, *Utopia* is essentially humanistic in outlook, for human qualities are valued and encouraged to develop to their maximum.

Under More's influence, the writing of imaginary societies began to flourish in the 17th century. Andreae's *Christianopolis* (1), published in 1619, describes a permissive, humanistic society organized around religion. Campanella's *City of the Sun* (5), published in 1623, is a community in which the state dominates the individual. Persons are regarded as servants of the state, and they must submit to regulation of occupational activity and private life. Bacon's *New Atlantis* (2), appearing in 1627, describes a society based upon science. Bacon sees the potentials of science as beneficial for mankind, and scientific knowledge is used for human betterment without being used for the control of individual behavior.

In the 18th and 19th centuries, utopian writing turned to socialism in the attempt to describe the ideal society. The 18th century generally is lacking in the production of major utopias, but Mercier's *Memoirs of the Year 2500* (11), appearing in 1770, is an appeal for establishing guidelines in society through a combination of science and religion. Spence's *Constitution of Spensonia* (19), Cabet's *Voyage to Icaria* (4), and Hertzka's *Freeland* (7), published respectively in 1803, 1840, and 1890, stress the goodness and equality of man, while proposing societies that reflect in one degree or another the socialistic conditions under which human nature is deemed to have its best chance for development. It is interesting to note that Cabet founded communities in the United States in an effort to realize his utopian dream, and that an attempt was made in Africa by followers of Hertzka to establish a society similar to that in *Freeland*. Bellamy's *Looking Backward* (3), appearing in 1888, is the most important of the socialistic utopias. It tries to preserve in-



dividual freedom while simultaneously featuring state control of the economy and elimination of class distinctions. *Looking Backward* reflects a much more humanistic brand of socialism than that, e.g., of Cabet.

The classical utopian tradition reached its climax with H. G. Wells' *Modern Utopia* (22). Published in 1905, this work expresses qualified optimism about the future of man and the possibility of using scientific methods of social organization to realize that future. However, it does not continue the tradition of expressing faith in man's goodness, rationality, and equality. Rather, the good future will be attained through centralization of control over the individual in the hands of a superior class of men (the samurai). The techniques used to accomplish this control make an interesting study of a set of proposals designed to revolutionize the social order.

#### D. ANTIUTOPIAS

The flavor of the antiutopias that have marked the writing of imaginary societies in the 20th century is seen clearly in Zamiatin's *We* (23). This work is preceded by other antiutopian writings, including Capek's *R.U.R.* (6), a story of a group of robots initially created to make life more utopian for man that turn on their masters and extinguish them. *We* directly anticipates *Brave New World* (8) and *1984* (15), the classical antiutopias, and offers a wealth of material regarding the behavior-control measures characteristic of a police state. Zamiatin's theme in *We* is that freedom and happiness are incompatible and that man will sacrifice the former for the latter. The society described operates with scientific precision to control the life of its citizens. Huxley's *Brave New World* (8) and Orwell's *1984* (15), appearing in 1932 and 1949 respectively, are so well known that no description of them is needed. It is sufficient to state that these works provide opportunity to examine the general topic of behavior control in considerable detail. Huxley's *Brave New World Revisited* (9), published in 1958, treats the issue of control in the light of recent scientific developments and concludes that the antiutopia described in his earlier work is well on its way toward realization. Finally, Khrushchev's *Documents of the 22nd Congress of the CPSU: Vol. 2, Report on the Program of the Communist Party of the Soviet Union* (10) provides a way to close this survey of imaginary societies with a reference to an existing society. Although individuals living in democracies normally classify communism as antiutopian, it is clear from Khrushchev's work that communists regard their views as utopian in the historical sense of the word, and these views can be analyzed and treated accordingly.

## E. CONCLUSION

It is probably true that the psychologist, as scientist, would be sympathetic to attempts to build utopias on empirical grounds. These attempts, after all, would be based on the belief that it is possible to have better conditions for humanity than now exist. This is an old belief that dies hard. However, the recent trend toward antiutopias may mean that this belief is dying and that pessimism, taking the form of a loss of faith in man's ability to use his knowledge for his betterment, is replacing it. Because the raw material for utopias is human nature, the psychologist should be helpful to, if not essential for, attempts to construct ideal societies. For this reason, it may be desirable for psychology and other branches of the life sciences to be applied to utopia building. Those who fear the results of using science in this way express themselves in antiutopias. Yet, the outcome need not be antiutopian in nature. Science may have the potential to contribute, either singly or in combination with other disciplines, to the creation of a better social order; to conclude at this stage of its development that it does not is surely not warranted.

## REFERENCES

1. ANDREAE, J. V. Christianopolis. In M. L. Berneri, *Journey Through Utopia*. Boston: Beacon Press, 1951. Pp. 34-57.
2. BACON, F. New Atlantis. In G. Negley & J. M. Patrick, *The Quest for Utopia*. Garden City: Doubleday, 1962. Pp. 353-372.
3. BELLAMY, E. Looking Backward. New York: New American Library, 1960.
4. CABET, E. A Voyage to Icaria. In G. Negley & J. M. Patrick, *The Quest for Utopia*. Garden City: Doubleday, 1962. Pp. 543-574.
5. CAMPANELLA, T. The City of the Sun. In G. Negley & J. M. Patrick, *The Quest for Utopia*. Garden City: Doubleday, 1962. Pp. 307-342.
6. CAPEK, K. R.U.R. In The Brothers Capek, *R.U.R. and The Insect Play*. London: Oxford, 1961. Pp. 1-104.
7. HERTZKA, T. Freeland. In G. Negley & J. M. Patrick, *The Quest for Utopia*. Garden City: Doubleday, 1962. Pp. 107-137.
8. HUXLEY, A. Brave New World. New York: Bantam Books, 1958.
9. ———. Brave New World Revisited. New York: Bantam Books, 1960.
10. KHRUSHCHEV, N. S. Documents of the 22nd Congress of the CPSU: Vol. 2, Report on the Program of the Communist Party of the Soviet Union. New York: Cross-currents Press, 1961.
11. MERCIER, L. S. Memoirs of the Year 2500. In G. Negley & J. M. Patrick, *The Quest for Utopia*. Garden City: Doubleday, 1962. Pp. 466-488.
12. MORE, T. More's Utopia and A Dialogue of Comfort. New York: Dutton, 1951.
13. MUMFORD, L. The Story of Utopias. New York: Viking Press, 1962.
14. NEGLEY, G. & PATRICK, J. M. The Quest for Utopia. Garden City: Doubleday, 1962.
15. ORWELL, G. 1984. New York: New American Library, 1961.



16. PLATO. The Republic. In E. H. Warmington & P. G. Rouse (Eds.), *Great Dialogues of Plato*. New York: New American Library, 1956. Pp. 118-422.
17. ROGERS, C. R., & SKINNER, B. F. Some issues concerning the control of human behavior. *Science*, 1956, **124**, 1057-1066.
18. SKINNER, B. F. *Walden Two*. New York: Macmillan, 1948.
19. SPENCE, T. The Constitution of Spensonia. In G. Negley & J. M. Patrick, *The Quest for Utopia*. Garden City: Doubleday, 1962. Pp. 489-502.
20. ST. AUGUSTINE. *The City of God*. New York: Dutton, 1931.
21. WALSH, C. *From Utopia to Nightmare*. New York: Harper & Row, 1962.
22. WELLS, H. G. A Modern Utopia. In G. Negley & J. M. Patrick, *The Quest for Utopia*. Garden City: Doubleday, 1962. Pp. 224-250.
23. ZAMIATIN, E. We. New York: Dutton, 1924.

*Department of Psychology*  
*Wake Forest College*  
*Winston-Salem, N. C.*

## THE RELATIONSHIP BETWEEN AN INDIVIDUAL'S SOCIOMETRIC STATUS IN DIFFERENT GROUPS OVER A TWO-YEAR PERIOD\*<sup>1</sup>

*U. S. Navy Medical Neuropsychiatric Research Unit, San Diego*

PAUL D. NELSON AND NEWELL H. BERRY

### A. INTRODUCTION

Sociometric methods of assessing individual performance have shown reliability and validity in many types of situations (11). Predictive validity of such measures has been reported for periods ranging from two to four years (2, 4, 6, 8). Longitudinal studies of peer-rating validity generally have utilized discrete on-the-job behaviors or superiors' evaluations as criteria. The question pursued in the present study is the extent to which peer evaluations obtained on an individual during the early stages of his membership in an organization are valid estimates of his later sociometric status among different peers in the same organization.

\* Experimental studies of reconstituted groups offer evidence for the stability of an individual's sociometric status over time, particularly on relatively general dimensions of behavior (1, 3, 10). The present study differs from the aforementioned studies in that a two-year period is covered, in contrast to periods from two to 16 weeks, and different rather than identical sociometric-criterion items were employed at the two times of assessment. At each assessment an attempt was made to evaluate the individual's status both in task-oriented and social-oriented domains, with the selection of specific sociometric items being based upon the assumed relevance of such items to the life situation at hand.

### B. METHOD

#### 1. Subjects

The Ss were 101 enlisted Marines who had completed two years of military service. The average age of Ss was close to 19 years; while the average

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 7, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This study was conducted as part of Research Project No. MR005.12-2201, (1) Report No. 65-5, under the Bureau of Medicine and Surgery, Department of the Navy. Opinions expressed are those of the authors and are not to be construed as necessarily reflecting the official view or endorsement of the Navy Department. The assistance of Michael McNally, HM1, USN is gratefully acknowledged.



level of formal education completed was approximately 11 years. Ss were part of a larger sample whose adjustment during the first-enlistment period was being studied and on whom a variety of data had been collected prior to and during recruit training two years previously. Ss were selected for the present study on the basis of their location, as a result of standard Marine Corps assignment procedures, in an accessible field battalion undergoing training in simulated combat operations.

## 2. Procedures

During the second and 10th weeks of basic training, the Ss and their recruit-platoon peers evaluated one another by nominating (*a*) those five men in the platoon who would make the best Marines and the five who would become the least desirable Marines and (*b*) the five men in the platoon they liked best and the five they liked least. The average size of the recruit platoons was close to 70 men during the second week of training and 65 men during the 10th week. Sociometric-status scores for Ss were derived separately on "best Marine" and "likability" items by summing algebraically the frequency of positive and negative nominations received and by dividing the sum by the factor  $\sqrt{N-1}$ , as suggested by Willingham and Ambler (13). *N* is the number of men making nominations.

At the two-year level, the Ss and peers from the seven field platoons in which Ss were located evaluated one another by nominating (*a*) the five men whom they considered most desirable to work with on day-to-day routine assignments and the five whom they considered least desirable to work with, (*b*) the five men whom they would most prefer as liberty companions and the five whom they would least prefer, and (*c*) the 10 men whom they would most prefer to have with them on a combat patrol behind enemy lines in war time. Because the Ss were within a restricted range of ranks after two years (pay graded E-1 through E-4) only those platoon members of the same range of ranks were given the sociometric questionnaire; and all nominations were restricted by instruction to members of the platoon of those particular rank levels, so that the raters and ratees would be peers in rank and experience to the maximum possible extent. The average number of platoon members providing nominations at the two-year level was approximately 28 men. Sociometric-status scores were derived on "work," "liberty," and "combat" items in the same way as was done in the second and 10th weeks of recruit training.

The membership of the recruit platoons from which the second-week and 10th-week sociometric nominations were obtained remained relatively stable

during the period of time involved and, in large part, the two-week and 10-week periods represented the limit of acquaintanceship among the men. On the other hand, the membership of the field platoons at the two-year level did not contain many men who had gone through recruit training together. Of the platoon members who evaluated Ss at the two-year level, an average of only two per cent had been in the same training platoons as the Ss two years previous. Furthermore, the nominations in training had been based upon two-week and 10-week periods of interpersonal contact; whereas the two-year nominations were based upon an average contact period of 10 months. Thus, in the field setting the men presumably were well acquainted with one another.

### C. RESULTS AND DISCUSSION

Relationships between sociometric-status scores obtained in recruit training and in different peer groups two years later were estimated by Pearson  $r$  coefficients. These values are shown in Table 1.

TABLE 1  
PEARSON CORRELATIONS BETWEEN INITIAL-TRAINING AND TWO-YEAR  
SOCIOMETRIC-STATUS VARIABLES  
( $N = 101$ )<sup>a</sup>

No.	Variable	Variable						
		1	2	3	4	5	6	7
1.	Two-week "best Marine"	.70						
2.	Two-week "likability"	.82	.65					
3.	10-week "best Marine"	.62	.73	.76				
4.	10-week "likability"	.46	.44	.44	.51			
5.	Two-year "work with"	.37 <sup>c</sup>	.31	.31	.35	.75		
6.	Two-year "liberty with"	.27	.27	.31	.36	.77	.61	
7.	Two-year "combat with"							

<sup>a</sup> All values are significant at the .01 level.

All the reliability coefficients for the "best Marine" and "likability" variables (as obtained through correlating second-week with 10th-week training scores) and all the validity coefficients of these recruit-training variables for predicting the two-year sociometric scores are within the range of values reported for comparable relationships in previous studies (11). Upon the basis of a test for differences between nonindependent correlations (9), no significant differences are found to exist between second-week and 10th-week variables in terms of their relationship with the two-year status variables; thus the data provide further support for the evidence obtained by Hollander (6) pertaining to the validity of early peer nominations.

Because there was some variation among the two-year platoon members



(a) in rank and (b) in the number of months in the platoon, the authors analyzed the effects of the attributes just named on two-year sociometric status. Months in the platoon was found not to be significantly related to any of the two-year status scores ( $p > .10$ ). Rank correlates .24 ( $p < .02$ ) with the work-status score but insignificantly with other scores. When rank is held constant, the partial correlations of recruit-training measures with two-year work status are found to be within one-half of a standard error of the original correlation values; thus the validity of sociometric status in training for predicting two-year status is not affected by rank at two years.

In terms of intercorrelation and relationships with subsequent criteria, the recruit-training scores "best Marine" and "likability" appear to have much in common. Further, the two-year variables "work" and "liberty" are highly correlated. Hollander (6) feels that communality such as the foregoing may be less a function of friends being chosen on task-oriented criteria than a function of persons with high status on task-oriented behavior being valued as friends. French holds a similar point of view. See his discussion of the relationship between liberty choices and sick-bay visits among Naval recruits (5).

Except for the difference between the two-year "work" and "liberty" variables in relation to the second-week "best Marine" variable, the correlations of recruit-training scores are significantly higher ( $p < .01$ ) with the two-year "work" variable than with the two-year "liberty" or two-year "combat" variable. One possible explanation for this result is that there exist different bases for the sociometric choices. The "best Marine" choices in recruit training and the two-year "work" choices are likely to entail consideration both of the individual's task competency and of his interpersonal effectiveness. Choices for "liberty," however, probably reflect general popularity as well as idiosyncratic factors.

Choices for "combat," especially from a group having little if any combat experience, may be based upon specific factors perceived by different group members as relevant but with little consensus among different individuals. Johnson, Waters, and Helme (7) for example, found that combat-potential ratings made during peace-time operations reflect underlying behavior factors different from those reflected by ratings made after combat experience.

Despite the possibility that different bases of judgment underlie the different sociometric choices, the two-year combat-status variable is more highly related to "work" ( $r = .77$ ) than to "liberty" ( $r = .61$ ). The difference between the correlations is significant at the .01 level. When the liberty-status variable is held constant, the partial correlation of "work" status with "combat" status remains relatively high (.70); but when the work-status variable is

held constant, the partial correlation of "liberty" status with "combat" status drops to .37. Because many of the factors underlying "work" status apparently are related to "combat" status, another possible explanation for the relatively low-order correlations of recruit-training variables with the two-year "combat" score is that the latter variable represents positive nominations only; whereas the remaining two-year status scores are based upon negative as well as upon positive nominations.

To check on the possible effect of score distribution on the correlations between the recruit-training and the two-year-status variables, the two-year "work" and "liberty" nominations were rescored, so that positive and negative nominations could be treated separately. From a practical point of view, this method of scoring is of relevance because, as Sells and Roff (12) have indicated, it is often difficult or unfeasible to have peers provide negative evaluations in operational field settings. The correlations involving the rescored "work" and "liberty" variables are given in Table 2.

When the work-status and liberty-status scores are based only on positive nominations, the correlations of the recruit-training variables with the two-year variables do not differ significantly from the correlations of recruit-training variables with two-year "combat" status.

TABLE 2  
PEARSON CORRELATIONS BETWEEN INITIAL-TRAINING AND TWO-YEAR POSITIVE AND  
NEGATIVE WORK-STATUS AND LIBERTY-STATUS VARIABLES  
( $N = 101$ )<sup>a</sup>

No.	Variable	Variable			
		5	6	7	8
1.	Two-week "best marine"	.34	.28	-.39	-.37
2.	Two-week "likability"	.28	.17	-.35	-.32
3.	10-week "best Marine"	.36	.28	-.23	-.20
4.	10-week "likability"	.39	.29	-.26	-.19
5.	Two-year positive "work"		.77	-.13	-.12
6.	Two-year positive "liberty"			-.16	-.17
7.	Two-year negative "work"				.92
8.	Two-year negative "liberty"				

<sup>a</sup> For  $N = 101$ ,  $.05 = .20$  and  $r .01 = .25$ .

Two further observations worth noting are as follows: First, as Sells and Roff (12) report, the positive-nomination scores and the negative-nomination scores on any one status variable are only moderately related. It also appears that negative nominees on the "work" and the "liberty" variables have more in common than do positive nominees. Second, the two-week peer nominations tend to correlate better than do the 10th-week nominations with the two-year negative-choice scores. Most striking are the differences between correlations



of the second and the 10th-week "best Marine" variable with negative choices on the "work" and the "liberty" variables, the differences being significant at the .01 level. It seems that very early peer nominations on a general dimension, such as "best Marine," are particularly sensitive to forms of behavior that at a later time are judged as dysfunctional or are negatively valued in the organizational setting.<sup>2</sup>

The close and continuous contact among individuals in basic-training settings seems necessary but insufficient to account for the validity of peer evaluations for predicting behavioral evaluations obtained much later. Basic military training, particularly for the adolescent, represents for many a traumatic break with the supports in life to which previously he had been accustomed. Response to such stress must be adaptive both to the demands of the organization and to the demands posed by the social context of his peer group. These demands confront the individual on the first day of training, but within two weeks rather stable estimates are formed among peers as to the capabilities that each man has in adjusting to novel and stressful situations. Once an individual has completed the requirements of training and has internalized the minimum values and goals of the organization of which he is a member, his on-the-job adjustment perhaps becomes less difficult. It seems that a person who has difficulty in adjusting to the early stresses in training tends to have difficulty in adjusting to unanticipated demands of subsequent work environments.

#### D. SUMMARY

The present study has demonstrated the extent to which an individual's sociometric status remains consistent over a two-year period even with membership changes in the peer group. Peer nominations for "best Marine" and "likability" obtained as early as the second week of basic military training are as valid as those obtained in the 10th week of training for predicting two-year "work," "liberty," and "combat" nominations. Second-week evaluations on the "best Marine" criterion are particularly sensitive to forms of behavior that after two years result in low sociometric status, as estimated from negative nominations.

#### REFERENCES

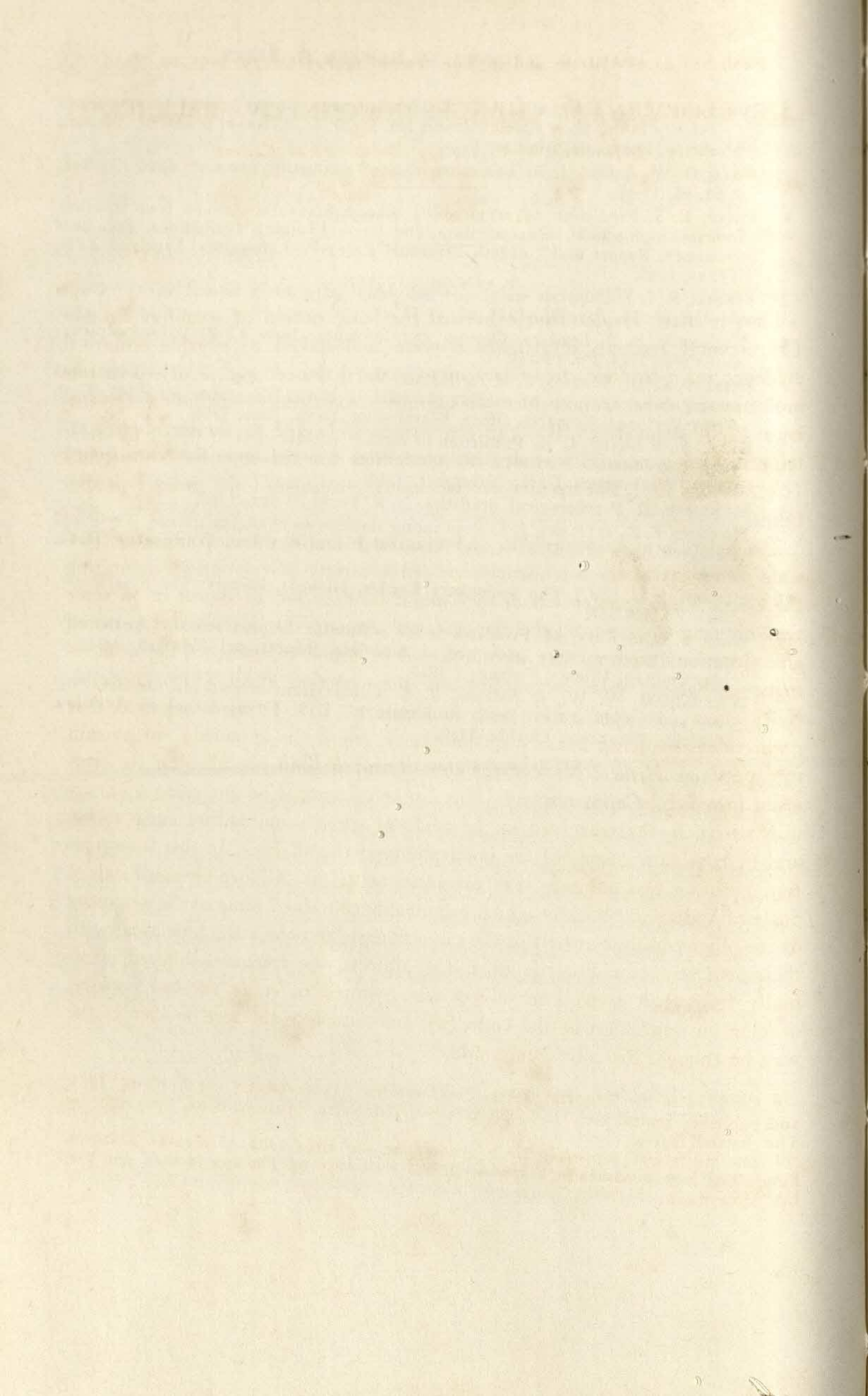
1. BELL, G. B., & FRENCH, R. L. Consistency of individual leadership position in small groups of varying membership. *J. Abn. & Soc. Psychol.*, 1950, **45**, 764-767.

<sup>2</sup> In another study in process, certain types of early peer evaluations are shown to be related to the criterion of premature separation from military service.

2. BERKSHIRE, J. R., & NELSON, P. D. Leadership peer ratings related to subsequent proficiency in the fleet. Special Report No. 58-20, U. S. Naval School of Aviation Medicine, Pensacola, Florida, 1958.
3. FISKE, D. W., & COX, J. A. The consistency of ratings by peers. *J. Appl. Psychol.*, 1960, **44**, 11-17.
4. FLYER, E. S. Prediction of unsuitability among first-term Airmen from aptitude indexes, high school reference data, and Basic Training evaluations. Technical Document Report 63-17, 6570th Personnel Research Laboratory, Lackland AFB, Texas, 1963.
5. FRENCH, R. L. Sociometric status and individual adjustment among Naval recruits. *J. Abn. & Soc. Psychol.*, 1951, **46**, 64-72.
6. HOLLANDER, E. P. Leaders, Groups, and Influence. New York: Oxford University Press, 1964. Pp. 85-138.
7. JOHNSON, C. D., WATERS, L. K., & HELME, W. H. Factor analysis of experimental noncognitive measures of combat potential. Technical Research Note 147, U. S. Army Personnel Research Office, Washington, D. C., June, 1964.
8. KATZ, A., & BURKE, L. K. Prediction of combat effectiveness of Officer Candidate School graduates. Technical Research Note 50, Personnel Research Branch, AGO, Washington, D. C., November, 1955.
9. MCNEMAR, Q. Psychological Statistics. New York: Wiley, 1962.
10. MEDLAND, F. F., & OLANS, J. L. Peer rating stability in changing groups. Technical Research Note 142, U. S. Army Personnel Research Office, Washington, D. C., 1964.
11. MORENO, J. L. (Ed.) The Sociometry<sup>\*</sup> Reader. Glencoe, Illinois: The Free Press, 1960. Pp. 320-387.
12. SELLS, S. B., & ROFF, M. Problems in the estimation of peer rejection in the elementary grades. Paper presented at American Educational Research Association, Chicago, Illinois, February, 1963.
13. WILLINGHAM, W. W., & AMBLER, R. K. Comparison of methods for deriving peer nomination scores. Research Report 35, U. S. Naval School of Aviation Medicine, Pensacola, Florida, 1962.

*U. S. Navy Medical Neuropsychiatric Research Unit  
San Diego 52, California*





## DEVELOPMENTAL CHANGES IN COGNITIVE BALANCE\*<sup>1</sup>

*Department of Psychology, University of Kansas*

---

ANTHONY J. SMITH

---

### A. INTRODUCTION

Shortly after Heider first expressed the basic notion of cognitive balance (5), several research investigations were undertaken to provide empirical evidence for a tendency to achieve or to prefer balance (e.g., 7, 8). The formulation of other conceptions of cognitive consistency—including congruence (11), symmetry (10), and consonance (4)—led to similar studies attempting to demonstrate equivalent tendencies toward cognitive consistency (e.g., 2, 11, 12). The results overwhelmingly confirmed the general predictions.

However, armchair analysis and empirical findings concerning the possible generality of these tendencies yielded a variety of restrictions delimiting the circumstances under which they might be expected to appear or in some way altering the form of their appearance. Among the various intrapersonal and situational factors that were assumed or found to influence the achievement of balance were incredulity and the assertion effect (11), the positivity effect (3), competition (6), and failure (16). It seems apparent that predictions involving balance or imbalance require a knowledge of certain intrapersonal characteristics of the subject and an understanding of the situation to which he must respond. This can be illustrated by analyzing a particular situation, that involved in the study of success and failure cited above, and by considering some of the implications of the analysis. In that investigation each of two subjects was presented with an Allport-Vernon-Lindzey Scale of Values test booklet which presumably contained some of the responses of the other person. For half of the subjects these responses were identical with their own responses. For the remaining subjects, the responses differed maximally from their own. The subject was required to study the test booklet, develop an impression of the colleague, and complete the test booklet in the way he thought his colleague would.

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 12, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The study was supported by a grant from the University of Kansas Research Fund. The author wishes to acknowledge the assistance of Thomas Jackson and Dr. Lee Silverthorn.



It has been proposed (14) that under these circumstances balance is achieved in the following manner. When, for example, the subject examines each answer of the hypothetical similar person, he finds that the response is like his own: e.g.,  $pLx$ ,  $oLx$ ; therefore in each instance balance tends to be achieved by the generation of a force in the direction of  $pLo$ . These forces summate to provide a positive attitude toward the other person. When the subject is required to predict the remaining answers of his colleague, the conditions that exist are  $pLo$  and  $pLx$ . This gives rise to a force in the direction of  $oLx$ . In brief, the perception of similarity in another person with respect to relevant matters produces acceptance. This acceptance in turn mediates the attribution of similarity.

This analysis suggests that if maximal balance is to occur (defined in terms of the objective stimuli) it will be necessary for the subject to be aware of and respond to each item differentially, so that each item contributes to the force toward acceptance.

In view of this observation, it seems reasonable to expect that the tendency to achieve cognitive balance in complex tasks might vary as a function of chronological age. The literature in developmental psychology contains many references to changes in differentiation associated with changes in age (e.g., 9, 19); for example, Werner (20) states that "wherever development occurs it proceeds from a state of relative globality and lack of differentiation to a state of increasing differentiation. . . ." This increasing differentiation occurs with respect to various kinds of psychological functions including perception and permits greater flexibility and effectiveness in relating to the external world. In the latter part of the life cycle, however, the process is reversed and dedifferentiation occurs (13, 17).

These considerations lead to the expectation that with increasing age subjects are more likely to respond differentially to the cues defining hypothetical similar and dissimilar individuals. On the basis of cognitive balance, one would predict that this differentiation would give rise to an increase in the differential acceptance of these hypothetical persons and an increase in the differential in the amount of similarity attributed to them (6, 14). Furthermore, this process should be reversed in later life: i.e., decreasing differentiation among the cues, with a resulting decrease in both differential acceptance and differential attribution of similarity. It is hypothesized, therefore, that the tendency toward balance increases from adolescence to maturity followed by a decline. These trends will be manifested (a) by an increase in the differential amount of similarity attributed to similar and dissimilar hypothet-

ical persons followed by a decrease among older subjects and (b) by an initial increase and a later decrease in the differential acceptance of these hypothetical persons. If acceptance is a crucial intermediate variable, the effects of balance as defined in terms of the objective situation will be more attenuated for the more remote differential attribution of similarity than for differential acceptance.

## B. METHOD

### 1. *Subjects*

Four groups of subjects totalling 189 were studied. The first group consisted of 17 Boy Scouts ranging in age from 11 through 15 with an average age of 13.32. Their cooperation in the research project had been elicited by the scoutmasters, and testing took place during regular scout meetings. The second group consisted of all of the students enrolled in the psychology courses in a high school. There were 91 such students (29 males and 62 females) with an age range from 16 through 19 and an average age of 17.63. The students were aware that they were subjects in an experiment related to their course work in psychology and were promised information concerning their own performance on the test instrument. All testing occurred during regular class periods. The third group consisted of 55 students (41 males and 14 females) enrolled in university courses in general psychology. Their ages ranged from 17 through 37 with a mean of 22.64. They participated in the experiment in order to fulfill one of their course requirements. This was essentially the same group reported in an earlier study (15). The final group, 26 members of domiciliary units in an installation of the Veterans Administration, ranged in age from 39 through 73 with a mean age of 57.73. They were selected in part on the basis of intellectual level. Their average AGCT score corresponded to the 70th percentile for national norms. Data were collected from them ostensibly as part of the regular station-testing program.

### 2. *Test Instrument*

The Allport-Vernon-Lindzey Scale of Values was given to all subjects before the experiment began. The Boy Scouts and high-school students were told that they were taking the test, so that they might become acquainted with it and be better prepared to participate in the subsequent research. The university students completed the test as part of their class work and no mention was made of it in relation to the project. Domiciliary members completed the test as part of an extensive battery that had just been introduced by the Psychology Service.



The test booklets were scored, and for each subject the value on which the highest score was obtained was designated as the critical value. Two experimental test booklets were then created for each of the subjects. Each booklet was partially completed, so that the responses that were inserted either duplicated those originally entered by the subjects, thereby creating a hypothetical similar person, or provided for the maximum systematic difference from the subject's responses resulting in an hypothetical dissimilar person. All information relating to the critical scale was eliminated from the experimental test booklets. The procedure employed in preparing these test booklets has been described in detail elsewhere (14).

### 3. *Instructions*

Several weeks after the administration of the Allport-Vernon-Lindzey Scale of Values the subjects were met in groups of varying size. They were told that in their everyday lives they were often obliged to estimate characteristics of other persons on the basis of limited information. They were about to be asked to do precisely that. They were to be given two test booklets that had been partially completed. Each was to be studied separately and, when the subject had developed a conception of the other person, he was to complete that test booklet in the way in which he felt the other person might have completed it.

The test booklets were then scored for the critical value.

### 4. *Measures of Attribution*

The degree of attribution of similarity was defined as the absolute difference between the subject's own score on the critical scale and the equivalent score from the experimental test booklet: i.e., the score attributed to the hypothetical person by the subject. Differential attribution of similarity was defined as the arithmetic difference between attribution of similarity scores for dissimilar and similar hypothetical persons; thus there could be no negative attribution of similarity scores, but negative differential attribution of similarity scores could arise if subjects attributed greater similarity to dissimilar others than to similar others.

### 5. *Measures of Acceptance*

Following the completion of the test booklets subjects were asked to rate on a graphic-rating scale the acceptability of each of the hypothetical persons as a work associate and as a leisure associate. These ratings of willingness to

associate with the other person were expressed in terms of the distance in millimeters from the negative end of the scale and could range from zero (very unwilling) to 125 (very willing). Ratings were not available for the university students, who had been studied independently several years earlier.

### C. ANALYSIS AND RESULTS

#### 1. Attribution

*a. The relationship between age and the differential attribution of similarity.* Attribution of similarity scores and differential attribution of similarity scores were determined for each subject, and means were computed for each of the four age groups. These means are shown in Table 1.

TABLE 1  
MEANS OF ATTRIBUTION OF SIMILARITY AND DIFFERENTIAL  
ATTRIBUTION OF SIMILARITY SCORES

Group	Characteristics of hypothetical other person		
	Similar	Dissimilar	Differential
Boy Scouts	8.53	9.18	.65
High-school students	7.48	13.88	6.40
University students	7.62	16.63	9.01
Domiciliary members	8.85	12.31	3.46

The variances of differential attribution of similarity scores were determined for each of the four groups and were tested for homogeneity by means of Bartlett's test. The resulting  $\chi^2$  is not significant ( $p > .25$ ). A simple analysis of variance was then performed ignoring the directional aspects of the hypothesis. The results of this analysis, as shown in Table 2(A) reveal significant differences among the age groups ( $p < .02$ ). The obtained

TABLE 2  
ANALYSES OF VARIANCE OF ATTRIBUTION-OF-SIMILARITY SCORES

Source	df	MS	F
A. Differential attribution of similarity			
Between groups	3	396.72	3.49*
Within groups	185	113.56	
B. Attribution of similarity for hypothetical similar persons			
Between groups	3	5.57	< 1
Within groups	185	39.45	
C. Attribution of similarity for hypothetical dissimilar persons			
Between groups	3	281.83	3.25**
Within groups	185	86.67	

\*  $p < .02$ .

\*\*  $p < .05$ .



pattern of means of differential attribution of similarity scores involves an increase from adolescence to maturity followed by a decrease among older subjects.

The differences between means for pairs of groups were tested, and significant differences were found for the following comparisons: Boy Scouts *vs.* high-school students ( $t = 2.69$ ,  $p < .01$ ), Boy Scouts *vs.* university students ( $t = 3.53$ ,  $p < .001$ ), university students *vs.* domiciliary members ( $t = 2.05$ ,  $p < .05$ ).

It also seemed desirable to determine the possible locus of the change associated with age. In the earlier study of the influence of success and failure (16) the author found that the effect occurred primarily in the attribution of similarity to dissimilar others, consequently separate analyses were undertaken of the relationship of age to the attribution of similarity to similar and to dissimilar persons.

*b. The relationship between age and the attribution of similarity to similar others.* The variances of the first set of attribution of similarity scores were computed and were determined to be homogeneous. An analysis of variance of these scores was then undertaken, and the results as summarized in Table 2(B) show no significant differences among the means.

*c. The relationship between age and the attribution of similarity to dissimilar others.* The variances of the second set of attribution of similarity scores involving dissimilar others were tested for homogeneity and were found to be homogeneous. The data were then analyzed by means of a simple analysis of variance. The means were found to differ significantly as shown in Table 2(C). The means of attribution of similarity scores involving dissimilar persons followed the same pattern found in the analysis of differential attribution of similarity scores. Once again the differences between the various pairs of means were tested and were found to be significant for these comparisons: Boy Scouts *vs.* high-school students ( $t = 2.78$ ,  $p < .01$ ), Boy Scouts *vs.* university students ( $t = 3.90$ ,  $p < .001$ ), university students *vs.* domiciliary members ( $t = 2.06$ ,  $p < .05$ ). These results also very closely resemble those found for differential attribution of similarity scores.

## 2. Acceptance

*a. The relationship between age and differential acceptance.* The mean ratings of similar and dissimilar persons and the mean differential ratings were determined for Boy Scout, high-school, and domiciliary subjects on each of the criteria. They are reported in Table 3.

TABLE 3  
MEANS OF ACCEPTABILITY RATINGS AND OF DIFFERENTIAL ACCEPTABILITY

Group	Similar	Criterion				
		Leisure Dissimilar	Differential	Similar	Work Dissimilar	Differential
Boy Scouts	83.12	66.35	16.77	96.47	72.24	24.23
High-school students	93.68	52.05	41.63	92.89	53.95	38.94
University students	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Domiciliary members	88.15	75.81	12.34	90.15	78.54	11.61

Variances of differential acceptance scores based on the criterion of leisure were computed for each of the three age groups and were tested for homogeneity. They were found not to be homogeneous, with the variance of the high-school group being significantly larger than the variances of each of the other groups. The differential acceptance scores were dichotomized at the median and entered in a two-by-three contingency table. A chi-square analysis reveals significant differences among the three groups ( $\chi^2 = 18.11$ ,  $p < .001$ ).

Comparisons among the three groups yield significant differences in two instances: the mean difference for high-school students is greater than those for Boy Scouts and domiciliary members ( $\chi^2 = 8.91$ ,  $p < .01$ ;  $\chi^2 = 32.33$ ,  $p < .001$ ).

Scores on the second criterion of acceptability (work) were analyzed in the same way. A test of homogeneity of variances once again yields significance, with the variance of domiciliary members being significantly lower than the variances of the other groups. Ratings were tallied in a two-by-three contingency table and analyzed by means of chi square. The resulting chi square of 8.68 is significant at the .02 level.

A test of the differences between groups yields significance in only one comparison: namely, the mean difference score for high-school students is significantly greater than that for domiciliary members ( $\chi^2 = 8.53$ ,  $p < .01$ ).

To determine the possible locus of this effect, separate analyses were undertaken for acceptance of similar persons and of dissimilar persons.

*b. The relationship between age and the acceptance of similar others.* The variances of ratings of the acceptance of similar others as leisure associates were determined for each of the three age groups. They were tested for homogeneity and were found to be heterogeneous. The variance for the domiciliary members is significantly lower than the variance for Boy Scouts or high-



school students. The scores were then dichotomized at the median, tallied in a two-by-three contingency table and tested for significance by means of chi square. The obtained chi square of 6.37 is significant at the .05 level. Differences between groups were tested and high-school students were found to be more accepting of the similar others than were the domiciliary members ( $\chi^2 = 6.37$ ,  $p < .02$ ).

Variances were computed for ratings based on the criterion of work associate and were found to be homogeneous. A simple analysis of variance was performed, with the results shown in Table 4(A). Differences among the means were found not to be significant.

TABLE 4  
ANALYSES OF VARIANCE OF RATINGS OF ACCEPTANCE

Source	df	MS	F
A. Acceptance of hypothetical similar persons as work associates			
Between groups	2	206	< 1
Within groups	130	604	
B. Acceptance of hypothetical dissimilar persons as leisure associates			
Between groups	2	6308	7.84*
Within groups	130	805	
C. Acceptance of hypothetical dissimilar persons as work associates			
Between groups	2	7337	8.48*
Within groups	130	865	

\*  $p < .001$ .

c. *The relationship between age and the acceptance of dissimilar others.* Variances of ratings of acceptance of dissimilar others as leisure associates were tested for homogeneity and the resulting chi square was found not to be significant ( $p > .10$ ). The ratings were then subjected to an analysis of variance, and differences among the means were found to be significant ( $p < .001$ ), as shown in Table 4(B).

The differences in mean ratings for each pair of groups were tested. The ratings of high-school students are significantly lower than those for Boy Scouts ( $t = 2.24$ ,  $p = .03$ ) and also lower than those for domiciliary members ( $t = 4.01$ ,  $p < .001$ ). The difference between Boy Scouts and domiciliary members is not significant.

Variances of ratings of acceptance of dissimilar others as work associates were also found to be homogeneous ( $p > .10$ ). The analysis of variance as summarized in Table 4(C) reveals significant differences among the groups ( $p < .001$ ).

It was found that high-school students had significantly lower mean rat-

ings than did either the Boy Scouts ( $t = 2.17$ ,  $p < .05$ ) or the domiciliary members ( $t = 4.29$ ,  $p < .001$ ).

#### D. DISCUSSION

In this investigation four groups were selected on the basis of chronological age to represent various stages of psychological development. No systematic effort was made to equate the groups. As a consequence, two differences emerged that could conceivably have affected the results: namely, the sex of the subjects and their intellectual level.

The two groups showing the highest differential attribution of similarity scores, high-school and university students, consisted of both male and female subjects; whereas the subjects in the remaining groups were male.

In an effort to determine the effect of sex, the difference between mean attribution of similarity scores for males and females were examined separately for high-school and university students. In neither instance is the difference significant ( $t = .78$ ,  $t = 1.30$  respectively), moreover, the differences are in opposite directions in the two groups. There is some reason to believe that the variable of sex did not confound the results.

Considering the complex nature of the task confronting the subjects, it would be possible to argue that the findings are the result of intellectual differences among the subjects—with the university students being a more highly selected group. Fortunately, scores on tests of intellectual ability were available for two of the groups and permitted an evaluation of this argument.

American Council of Education scores were obtained for 44 of the university students who had taken the test as part of an entrance examination required of all students entering the University. The correlation between the students' total ACE scores and their differential attribution of similarity scores was found to be insignificant ( $r = +.24$ ,  $p > .10$ ).

AGCT scores were available for all of the domiciliary members. For this group, the correlation between AGCT score and differential attribution of similarity was also found to be insignificant ( $r = -.11$ ,  $p = .70$ ). These results make it seem less likely that the present findings can in any way be attributed to differences in intellectual ability among the groups.

It seems reasonable to conclude that the general findings provide support for the hypotheses under investigation. With increasing age there is an increasing differentiation between the hypothetical similar other and dissimilar other, with a reversal of this trend in the highest-age group. The curvilinear trend holds for differential attribution of similarity and for dif-



ferential acceptance on the basis of each of the two criteria: work and leisure. These results are due primarily to differences in the acceptance of and attribution of similarity to hypothetical dissimilar others. With the exception of acceptance based on a criterion of leisure, there are no differences in the treatment of hypothetical similar persons that are associated with age.

It is of interest to note the relationship between the present findings and those reported in earlier studies. The pattern of responses that is characteristic of high-school students and domiciliary members as presented in Table 1 closely resembles that reported earlier for subjects who had experienced failure (16). In that study the mean attribution of similarity scores for university students who had previously failed in a series of tasks was 7.17 for similar others and 12.61 for dissimilar others. Moreover, the results for the present university students are not only similar to those in another earlier study (15), which reported values of 7.05 and 16.18, but approximate fairly closely those found for subjects who had experienced success (8.38 and 17.94). This suggests the possibility that the differences obtained in the study of the influence of success and failure might have been the result of regression induced by the frustrating experience of failure (1).

It had been noted earlier that if attribution of similarity were mediated by acceptance, the subjects should be more discriminating with respect to their ratings of the acceptability of the similar and dissimilar others than they are with respect to the attribution of similarity. It was possible to make comparisons for three groups for whom both sets of data were available: Boy Scouts, high-school students and domiciliary members. The relevant data can be found in Tables 1 and 3.

Among the Boy Scouts the difference in the mean acceptability ratings for similar and dissimilar others was significant for both criteria: leisure ( $t = 2.16$ ) and work ( $t = 2.38$ ). The difference in mean attribution of similarity ratings was not significant ( $t = .36$ ).

The results obtained for high-school students (leisure,  $t = 8.83$ ; work,  $t = 7.74$ ; attribution of similarity,  $t = 5.88$ ) and for domiciliary members (leisure,  $t = 2.23$ ; work,  $t = 2.02$ ; attribution of similarity,  $t = 1.47$ ) were in the same direction.

In each instance the results are consistent with the proposal that the effects of balance as defined in terms of the objective situation are more attenuated for the more remote differential attribution of similarity than for differential acceptance.

The results of the present study appear to be in conflict with those de-

scribed by Weir (18), who found that balance occurs less frequently among high-level than among low-level university students. Although a direct comparison between the two studies is virtually impossible because of differences in the sample, the task, and the operational definition of balance, it seems that the apparent conflict might arise primarily because of the differing nature of the two tasks. Weir's fact-value differentiation technique yields balance when subjects in responding to individual statements (e.g., "University Teachers are mostly comparatively free from prejudice") distort their judgments of the factual nature of the statements to conform to their own evaluation of the topic to which the statement refers: i.e., university teachers. The more mature subjects tend to maintain a clearer distinction between their evaluative and factual judgments.

Weir's use of statements that can be judged factually and his task in which balance is achieved by a distortion of fact distinguishes his technique from the one employed in the present study, in which the attainment of balance does not involve distortion and there is no prior basis for evaluating the accuracy of judgments. This underlines the fact that any generalizations that are made concerning developmental changes in cognitive balance require the specification of the conditions involved in the task.

\* In the present study cognitive balance appears to be maximized when the subject responds differentially to each of the elements in a complex task and when there is no restriction imposed on his judgments by reality requirements.

#### E. SUMMARY

Four groups ranging in average age from 13.32 to 57.73 were selected to represent various stages of psychological development. Each *S* was given two partially completed Allport-Vernon-Lindzey Scale of Values booklets constructed to represent similar and dissimilar individuals. *S* studied each booklet and attempted to reproduce the remaining answers. Ratings were obtained of each of the hypothetical persons on two criteria: leisure associate and work associate. It was found that with increasing age there was an increasing differential in the acceptance of similar and dissimilar others and an increase in the differential attribution of similarity. The trend was reversed for older subjects. The results are consistent with an interpretation based on cognitive balance.

#### REFERENCES

1. BARKER, R. G., DEMBO, T., & LEWIN, K. Frustration and regression: An experiment with young children. *Univ. Iowa Stud. Child Welfare*, 1941, 18.



2. BREHM, J. W. Postdecision changes in the desirability of alternatives. *J. Abn. & Soc. Psychol.*, 1956, **52**, 384-389.
3. DEUTSCH, M., & SOLOMON, L. Reactions to evaluations by others as influenced by self evaluations. *Sociometry*, 1959, **22**, 93-112.
4. FESTINGER, L. A Theory of Cognitive Dissonance. Evanston: Row, Peterson, 1957.
5. HEIDER, F. Attitudes and cognitive organization. *J. of Psychol.*, 1946, **21**, 107-112.
6. ———. The Psychology of Interpersonal Relations. New York: Wiley, 1958.
7. HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4**, 57-76.
8. JORDAN, N. Behavioral forces that are a function of attitudes and of cognitive organization. *Hum. Relat.*, 1953, **6**, 273-287.
9. LEWIN, K. Principles of Topological Psychology. New York: McGraw-Hill, 1936.
10. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393-404.
11. OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42-55.
12. RUNKEL, P. J. Cognitive similarity in facilitating communication. *Sociometry*, 1956, **19**, 178-191.
13. SCHAEIE, K. W. A field-theory approach to age change in cognitive behavior. *Vita humana, Basel*, 1962, **5**, 129-141.
14. SMITH, A. J. Similarity of values and its relation to acceptance and the projection of similarity. *J. of Psychol.*, 1957, **43**, 251-260.
15. ———. Perceived similarity and the perception of similarity: The influence of valence. *J. Abn. & Soc. Psychol.*, 1958, **57**, 376-379.
16. ———. The attribution of similarity: The influence of success and failure. *J. Abn. & Soc. Psychol.*, 1960, **61**, 419-423.
17. WAPNER, S., WERNER, H., & COMALLI, P. Perception of part-whole relationships in middle and old age. *J. Gerontol.*, 1960, **15**, 412-416.
18. WEIR, A. J. A developmental measure of fact-value differentiation. *Vita humana, Basel*, 1960, **3**, 65-82.
19. WERNER, H. Comparative Psychology of Mental Development. New York: Harper, 1940.
20. ———. The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The Concept of Development*. Minneapolis: Univ. Minn. Press, 1957. Pp. 125-148.

Department of Psychology

University of Kansas

Lawrence, Kansas 66045

## INCIDENTAL STIMULI: SOME EFFECTS OF COLOR AND MAGNITUDE IN A CONCEPTUAL SITUATION\*

*Department of Psychology, Florida State University*

DAVID B. COHEN AND WILLARD H. NELSON

### A. INTRODUCTION

Early research indicated that incidental learning occurs, but that it is much less efficient than intentional learning (1). Later studies have indicated that the amount of incidental learning is related to the duration of the incidental stimuli (2) and also is dependent on the number of incidental-stimulus presentations (6). Other studies indicate that the amount of incidental learning is dependent on the type of orienting task (4, 5). Further, incidental learning can be expected to vary with the ability to discriminate and categorize stimulus materials along more than one dimension (3). Results for certain color combinations approach significance at intensities close to the statistical threshold (7), while Nelson and Cohen (in unpublished research) have obtained no significant effects on learning by ninth grade Ss who were presented differently colored incidental stimuli. Differences in procedure, as well as age range, suggests possible sources of difference in results.

### B. METHOD

#### 1. *Subjects*

The Ss were 10 male and 10 female Florida State University students enrolled in sophomore-level psychology classes.

#### 2. *Apparatus and Materials*

The apparatus consisted of two 35 mm projectors, a powerstat, a tachistoscopic device, and a projection screen.

The materials included 12 focal slides, each showing in different order a triangle, a square, an octagon, and a circle. Four slides were used to produce the incidental stimuli. The cue slide for a triangle showed the word "three," for a square the word "four," for an octagon the word "eight," and for a circle the word "round." Two Kodak Wratten gelatin filters were employed: a No. 12, yellow, and a No. 25a, red.

\* Received in the Provincetown Editorial Office on April 12, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



### 3. Procedure

Each *S* was treated individually and was led to believe that the experiment concerned ESP. *Ss* were instructed with respect to how and when they were to respond to the focal material to be shown.

The objective of the first part of the study was to determine the incidental stimulus magnitude at which *Ss* make maximum use of the cues, as indicated by the making of 15 or 16 correct choices out of 16. A series of 16 choices was termed a trial, and the number of trials varied according to the rate of reaching criterion. All trials used the 300-watt setting to project the focal material. The projectors were 15 feet and the *S* was 12 feet from the screen. The incidental cues were projected 10 times in a 15-second interval for each focal slide, with intensity and duration systematically varied. Both intensity and duration were varied to get one composite indicator (i.e., the magnitude) of these two foregoing factors.

The magnitude on the first trial was .2 volt second; that is, a duration of .01 second at 20 volts for the incidental cue. Successive trials were made at 10-volt increments up to 120 volts at the .01-second duration. For those *Ss* who had not used the incidental cues at 1.2 volt seconds, brief trials were made at successively increasing voltages and durations up to the limits of 120 volts. At the magnitudes above 1.2 volt seconds, the cue words were clearly discernible for the *Ss* tested (as was determined by questioning following the completion of testing).

Some *Ss* so consistently gave wrong answers that apparently they could only be doing so by processing the cues in some way; hence when such *Ss* had made 15 or 16 *wrong* choices they were also considered to have reached criterion. Questioning of the *Ss* at the end of testing verified the appropriateness of treating their data with those of *Ss* making 15 or 16 *correct* choices.

The objective of the second part of the experiment was to determine whether or not differences of color of incidental stimuli and focal slides were systematically related to the *S's* choice behavior. Each *S* was given two trials of 16 selections each. Colored filters were used to produce yellow focal geometric figures and red incidental-cue words for the first trial, and red focal geometric figures and yellow incidental-cue words for the second trial. [Vinacke (7) obtained his largest differences with these colors]. The intensity used to project the focal material was the 300-watt setting. The magnitude for the incidental-cue word was equivalent to that at which each *S* (with one exception in which the *S* demonstrated unusual cue use) obtained

a greater-than-chance number of correct choices with noncolored geometric figures and cues.

### C. RESULTS AND DISCUSSION

Ten males and four females made maximum use of the incidental stimuli; seven males and four females made 15 or 16 correct choices out of 16 selections, and three males made one or no correct choices out of 16 selections. The number of males using the cues was greater than the number of females, the difference being significant at the .02 level. There is a significant difference at the .05 level between the number of Ss using and not using the cues. For those Ss who used the cues, the mean magnitude for the males was .71 volt seconds; for the females, .55 volt seconds ( $t = 2.22$ ,  $p < .05$ ).

For all Ss, the relation between the number of correct choices and the several magnitudes of noncolored cues is significant at the .05 level ( $F = 6.80$ ), but there is no significant sex difference ( $F = .04$ ) nor a significant sex-by-magnitude interaction ( $F = .24$ ).

The mean magnitude of incidental colored cues for males is .48 volt seconds; that for females, .39 volt seconds ( $t = 3.58$ ,  $p < .01$ ). An analysis of variance of correct choices for the red-yellow and yellow-red pairings of color of focal stimuli and of incidental cues yields nonsignificant  $F$ s of 1.4 for color and 1.2 for interaction. For sex the  $F$  is 17.2 ( $p < .01$ ).

The results of the study imply that learning a conceptual task of the type employed is facilitated by either noncolored or colored incidental-word stimuli, with the tentative provision that with colored materials only females are likely to demonstrate significantly improved learning. A further implication is that the use of incidental noncolored stimuli, when presented at gradually increasing magnitudes, may be related to personality traits because those Ss who did not use the stimuli to the maximum or who used them to the maximum in a negative manner stated that they felt their action was a demonstration of independence.

### REFERENCES

1. MCGEOCH, J. A., & IRION, A. L. *The Psychology of Human Learning*. New York: Longmans, 1952. Pp. 210-215.
2. NEWMARK, E., & SALTZMAN, I. J. Intentional and incidental learning with different rates of stimulus presentation. *Amer. J. Psychol.*, 1953, **66**, 618-621.
3. PENDERLEITH, M., & POSTMAN, L. Discrimination and verbal habits in incidental learning. *Amer. J. Psychol.*, 1956, **69**, 236-243.
4. POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: IV. The interaction of orienting tasks and stimulus materials. *J. Exper. Psychol.*, 1956, **51**, 329-333.



5. SALTZMAN, I. J. Comparisons of incidental and intentional learning with different orienting tasks. *Amer. J. Psychol.*, 1956, **69**, 274-277.
6. SALTZMAN, I. J., & ATKINSON, R. L. Comparisons of incidental and intentional learning after different numbers of stimulus presentation. *Amer. J. Psychol.*, 1954, **67**, 521-524.
7. VINACKE, W. E. The discrimination of color and form at levels of illumination below conscious awareness. *Arch. Psychol.*, 1942, **38**, 5-53.

*Department of Psychology*  
*Florida State University*  
*Tallahassee 32306*

# TYPE OF REWARD, INCENTIVE, AND INCENTIVE-SEQUENCE AS FACTORS IN THE MOTOR PERFORMANCE OF MENTALLY RETARDED, PHYSICALLY HANDI- CAPPED, AND COLLEGE STUDENTS\*<sup>1</sup>

*New England Rehabilitation Research and Demonstration Institute,  
Northeastern University*

SALLY L. PERRY AND BERNARD A. STOTSKY

## A. INTRODUCTION

The effectiveness of various types of rewards and incentives in motivating performance of the mentally retarded has been studied by several investigators. Gordon, O'Connor, and Tizard (4) found that a group of retarded given a specific goal to strive for (goal group) performed significantly better on a motor task than did either a group divided into two teams, each team encouraged to beat the other (cooperation group); or a group divided into pairs, each pair member urged to beat his partner (competition group). All three incentive groups performed significantly better than did a control group. In comparing verbal incentives, Stevenson and Snyder (6) found that neutral and positive incentives were more effective in motivating the motor performance of the retarded than were negative incentives; moreover it was found that performance varied as a function of the sequence of incentives.

Hunt and Patterson (5) rewarded two groups of retarded children with candy or verbal urging for repetitive performance on the Goodenough Draw-A-Man test. No overall difference between reward types was found. Using a rotary-pursuit motor task, Ellis and Distefano (2) found that verbal urging and praise served to increase the performance of a group of retarded significantly above that of a control group. In a reaction-time experiment, Wolfensberger (7) found no differences among groups concretely rewarded, symbolically rewarded, concretely punished, and symbolically punished and a control group. Toys served as the concrete reward, and tokens to be ex-

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 14, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This study was supported in part by the Vocational Rehabilitation Administration (Grant No. RD-1310-G).

Acknowledgments are made to Dr. George J. Goldin and Dr. Reuben J. Margolin, the Misses Ann Donovan, Dorothy Thorne, and Elizabeth Russell, and William Twomey for assistance. Special appreciation is extended to Mrs. H. Scott Meyers, William J. Carmichael, and D. Salvi for making the research possible.



changed for toys served as the symbolic reward. Punishment was defined as the taking away of toys or tokens. Zigler and Unell (8) compared two groups of retarded children on a concept-sorting task. A group tangibly reinforced with toys switched more readily from one sorting principle to another than did a group receiving no reward.

In only one study cited were specific incentives provided to the subjects prior to the performance of the task (4). The remaining studies dealt with the effect of reward or punishment itself on performance. The rewards employed were material (candy or toys) or verbal. In attempting to bring about behavioral changes it may be important to distinguish between incentives, which affect the motivational state of the subject in approaching the task (either in a positive or negative direction), and reinforcement, which affects the probability of the occurrence of the response. (If positive, reinforcement increases the probability of the occurrence of the response; if negative, it decreases the probability of the occurrence of the response.)

In considering incentives, much behavior is motivated by the desire to avoid failure, disapproval, or loss of some material object as well as by the desire for success, praise, and acquisition of some material object. Reward, in turn, may affect the individual's performance (*a*) by strengthening a desire for some object or by strengthening a determination to avoid an unpleasant state of affairs, such as loss of reward or punishment; or (*b*) by reducing desire once the reward is achieved. Therefore, one may very well expect an interaction between incentive and reward in the experimental situation. The incentive would be embodied in the instructions prior to the performance of the experimental task, while reward would be the consequence of performance meeting the criterion contained in the instructions. The consideration of the differential aspects of incentives and rewards is important in the sense that the effect of incentive on the motivational state of the individual may influence response to reward. On the other hand, the type of reward (as well as the cumulative amount of reward) may result in significant changes in the motivational state of the subject, regardless of incentive.

A variable considered in only one of the studies cited is the effect of the sequence of incentives on performance (6). For example, would subjects given a positive incentive first (an incentive to perform well with no threat of punishment or reproof) or a mixed incentive (an incentive to perform well and to avoid punishment or reproof) perform differently from subjects given a negative or mixed incentive first followed by a positive incentive? In the experiment to be reported the effect of incentive sequence was limited to material rewards and nonverbal nonmaterial rewards.

The purposes were (a) to investigate the relative effects of the type of reward, the type of incentive, and the sequence of incentives on the motor performance of a group of mentally retarded adolescents; and (b) to compare the performance of physically handicapped adolescents and adult controls with the mentally retarded as to the differential effects of these variables. The variables studied were (a) material and nonmaterial reward, (b) positive and mixed incentives, and (c) the sequence of incentives. It was expected that in each population (regardless of the incentive sequence) the material and the nonmaterial rewards would increase the performance level of the experimental groups over that of the control group. Further, it was anticipated that material reward would be more motivating than nonmaterial reward to the mentally retarded; and that nonmaterial reward would be more motivating than material reward to the adults. No prediction was made as to the effect of type of reward on the performance of the physically handicapped nor were assumptions made concerning the effects of the type of incentive or the sequence of incentive on performance.

## B. METHODOLOGY

### 1. Design

The experimental design was a two-by-two-by-two factorial in which type of reward (material or nonmaterial), type of incentive (positive or mixed), and sequence of incentive (positive followed by mixed or mixed followed by positive) represented the dimensions. The five groups in each of the three populations were (a) material reward with positive incentive followed by mixed incentive, (b) material reward with mixed incentive followed by positive incentive, (c) nonmaterial reward with positive incentive followed by mixed incentive, (d) nonmaterial reward with mixed incentive followed by positive incentive, and (e) a control group. A total of 13 trials were given: three control trials; and five experimental trials under each of the two incentive conditions. The control group performed the same task for all 13 trials, but did not receive the experimental treatments.

### 2. Subjects

Fifty mentally retarded persons from two vocational-training centers, 43 orthopedic-neuromuscular students from an industrial school for crippled children, and 35 college-graduate students in education served as subjects (Ss). Within each of the three populations sampled, the five groups were matched (when possible) for background factors such as age, intelligence,



and educational achievement. For the mentally retarded population, age varied from 16 to 42 with a mean of 20; *IQ* varied from 52 to 89 with a mean of 68. For the orthopedic-neuromuscular population, age varied from 12 to 20 with a mean of 16; *IQ* varied from 73 to 121 with a mean of 95. Of the 43 orthopedic-neuromuscular students, 14 were polio cases; 11, spina bifida; 11, cerebral palsy; three, muscular dystrophy; one, epilepsy; one, heart; one, amputee; and one, congenital deafness. For the graduate-student population, age varied from 22 to 58 with a mean of 34; *IQ* was not available, but all had completed college and had the equivalent of approximately one year of graduate work in education. No subject had a gross sensory or motor defect.

### 3. *Materials and Procedure*

The materials for each subject consisted of (a) 16  $2\frac{1}{2}$ " bolts, eight of which were  $\frac{3}{16}$ " in diameter and eight of which were  $\frac{1}{4}$ " in diameter; (b) 16 nuts corresponding in diameter to the diameters of the bolts; (c) 32  $1" \times 1" \times \frac{3}{4}"$  wooden blocks, with a hole in the center of each block corresponding in diameter to the bolt diameters; and (d) a  $9\frac{3}{4}" \times 9\frac{3}{4}" \times \frac{3}{4}"$  wooden board containing 16 holes (arranged in four rows of four holes each). The subject's task was to assemble the materials by bolting two blocks, one to each side of the board, and tightening a nut to the bolt. Each such unit was called a double block.

Prior to the introduction of rewards, three two-minute control trials were given during which time *S* was instructed to assemble as many double blocks as he could within the allotted time period. Following the control trials, the system of rewards was explained to each of the four experimental groups; then the 10 experimental trials were given. The control group was merely instructed to continue as before. The intertrial interval was six minutes. During this time the experimenter (*E*) recorded the number of double blocks completed and passed out rewards, while *Ss* disassembled the double blocks and rested.

The control trials on which *S* assembled the greatest number of double blocks served as the baseline upon which subsequent reward was allotted. Under the positive incentive condition, *Ss* were rewarded for each trial performance better than that of their top control-trial performance; for failure to improve, no reward was given. Under the mixed-incentive condition, *Ss* were rewarded for each trial performance better than that of their top control-trial performance; but for failure to improve, reward was taken away.

Material reward for all populations consisted of increases or decreases in the number of nickels. Money was chosen as the material reward because money, as a generalized reinforcer, was assumed to meet whatever material-deprivation need existed in *S*. For the mentally retarded population, the nonmaterial reward consisted of increases or decreases within a rating system similar to the rating systems used at the training centers; for the orthopedic-neuromuscular and the graduate-student populations, the nonmaterial reward consisted of increases or decreases within a letter-grade system similar to the letter-grade systems used at the industrial school and at the graduate school, each system paralleling in the number of categories the rating system used for the mentally retarded. At the outset of the experimental trials, all materially rewarded *Ss* were given five nickels; nonmaterially rewarded mentally retarded *Ss* were given the rating "Fair," and the nonmaterially rewarded orthopedic-neuromuscular and graduate-student *Ss* were given the letter grade "C." In the positive-incentive condition, *S* was given a nickel or one classification increase in letter grade or rating for each double block assembled over and above the number assembled on his top control trial. In the mixed-incentive condition, *S* was rewarded in the same fashion as in the positive-incentive condition for performance better than the top control trial; however when performance was equal to or below that of the top control trial *S* lost a nickel or was dropped one letter grade or rating classification. Placed in front of each *S* was a sheet of paper containing his name and a horizontal row of 11 squares. Above each square was written, from left to right, 0¢, 5¢, 10¢, 15¢, 20¢, 25¢, 30¢, 35¢, 40¢, 45¢, and 50¢ for materially rewarded *Ss*; E—, E, D—, D, C—, C, C+, B, B+, A, and A+ for nonmaterially rewarded orthopedic-neuromuscular and graduate student *Ss*; and Very Unsatisfactory, Unsatisfactory, Very Poor, Poor, Not Quite Poor, Fair, Not Quite Good, Good, Very Good, Not Quite Excellent, and Excellent for nonmaterially rewarded mentally retarded *Ss*. After each trial, *E* recorded below the squares the number of double blocks completed. Noting *S*'s performance, *E* then placed in or removed from the squares the appropriate number of nickels or colored or erased the appropriate number of squares.

### C. RESULTS

Figures 1, 2, and 3 present the number of double blocks completed on each two-minute control trial and experimental trial for the mentally retarded, orthopedic-neuromuscular, and graduate-student populations respectively.



An analysis of variance for the three control trials for each of the three populations indicated no differences among the groups. Because the groups did not differ during the control period, an analysis of variance was performed for the 10 experimental trials to determine the effect due to experimental treatments. No overall differences were obtained with respect to type of reward, type of incentive, or sequence of incentive; and there were no

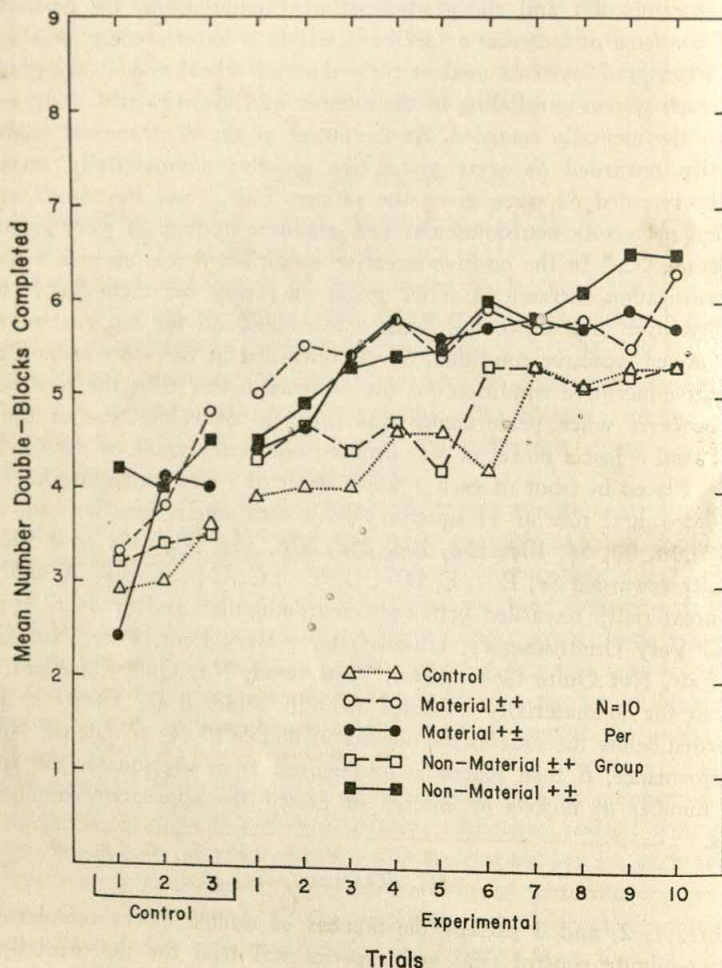


FIGURE 1

THE MEAN NUMBER OF DOUBLE BLOCKS COMPLETED BY THE MENTALLY RETARDED ON EACH OF THE THREE CONTROL AND 10 EXPERIMENTAL TRIALS

significant interactions. The main effect due to trials was significant for all populations: for the mentally retarded,  $F = 39.72$  ( $df = 12, 576$ ;  $p < .001$ ); for the orthopedic-neuromuscular students,  $F = 50.60$  ( $df = 12, 492$ ;  $p < .001$ ); and for the graduate students,  $F = 48.65$  ( $df = 12, 396$ ;  $p < .001$ ).

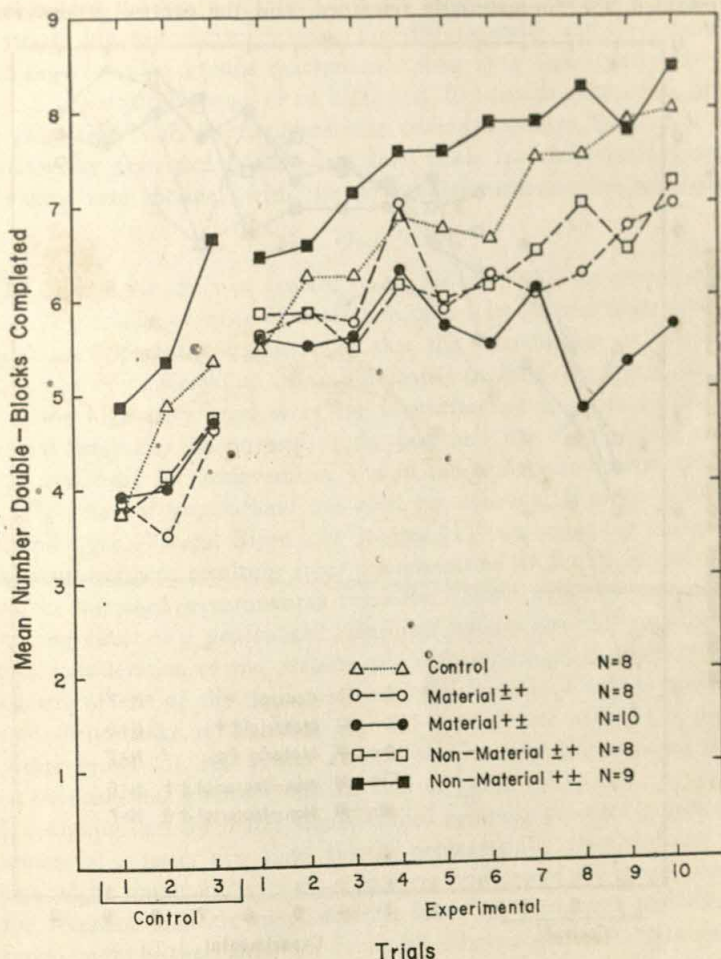


FIGURE 2  
THE MEAN NUMBER OF DOUBLE BLOCKS COMPLETED BY THE ORTHOPEDIC-  
NEUROMUSCULAR ON EACH OF THE THREE CONTROL AND 10  
EXPERIMENTAL TRIALS



From preliminary trials it was evident that the task imposed a ceiling on performance; subsequently it was found that nearly half of the *Ss*, regardless of population, had reached their top-performance level by the fifth experimental trial. In addition, the intrasubject range of performance was found to be minimal. A *t* test was then applied to the data to compare the materially rewarded, the nonmaterially rewarded, and the control groups for the

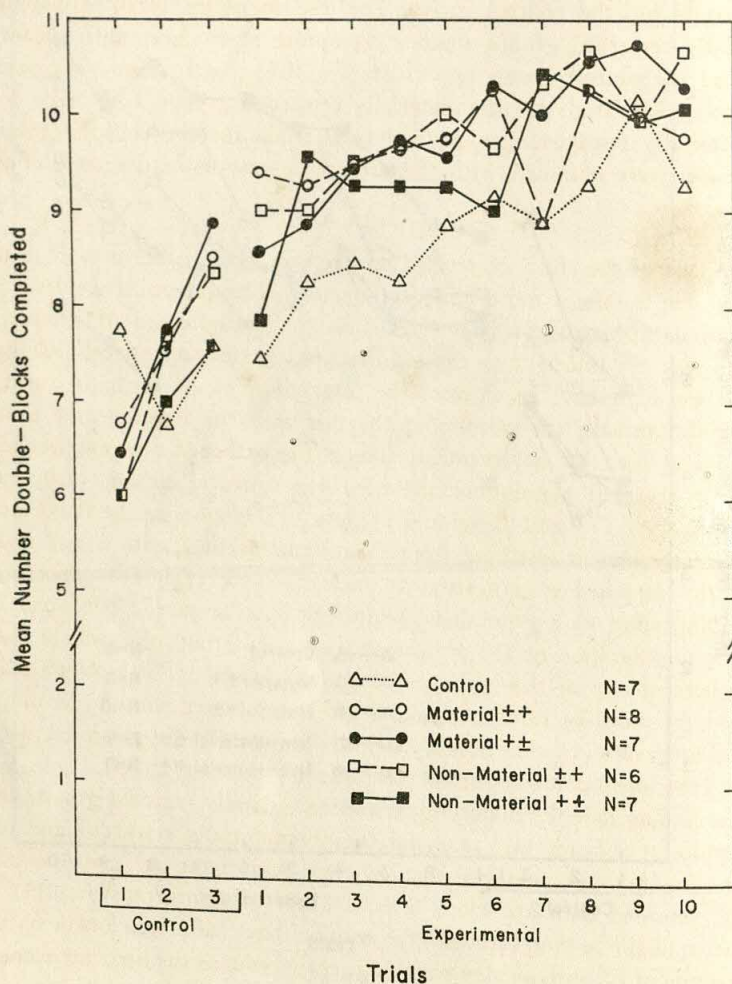


FIGURE 3

THE MEAN NUMBER OF DOUBLE BLOCKS COMPLETED BY THE GRADUATE STUDENTS ON EACH OF THE THREE CONTROL AND 10 EXPERIMENTAL TRIALS

number of trials to peak performance. Peak performance was defined as the first experimental trial on which *S* reached his best performance. In the graduate-student population, the nonmaterially rewarded groups reached asymptote at a significantly faster rate than did the control group ( $t = 1.87$ ,  $p < .05$ ): with mean peak performance being 6.00 trials for the nonmaterially rewarded groups, 6.70 trials for the materially rewarded groups, and 8.00 trials for the control group. For the mentally retarded population, materially rewarded groups reached asymptote at a significantly faster rate than did the control group ( $t = 1.93$ ,  $p < .05$ ): with mean peak performance being 4.80 trials for the materially rewarded groups, 6.40 trials for the nonmaterially rewarded groups, and 6.70 trials for the control group. No differences were obtained with the orthopedic-neuromuscular population.

#### D. DISCUSSION

The lack of the effect of reward on performance (in the overall analysis) may be due to one or more of several factors. The general drive level in all populations apparently was so high that the introduction of specific reinforcers was not sufficient to cause differential performance levels. Contributing to the high-drive level were the interaction of *Es* with *Ss* during the intertrial intervals, the novelty of the task and the experimental situation, the general need for achievement, and in the orthopedic-neuromuscular and mentally retarded populations the need for approval through achievement. Wölffensberger (7) and Bijou and Sturges (1) emphasize the importance of social reinforcement resulting from the interaction of *S* with *E* and its effect upon the supposed experimental reinforcer. Some evidence supporting the overriding effect of a generalized reinforcer upon a specific reinforcer comes from a consideration of the performance of a subgroup of graduate students (members of one of the graduate education classes). These *Ss* verbally expressed their dislike for the task and did not interact with *Es* as did *Ss* in the other groups. In this group, reward significantly differentiated the combined experimental groups from the control group ( $p < .05$ ).

The finding that the materially rewarded mentally retarded groups reached asymptote at a faster rate than did the nonmaterially rewarded and control groups might simply indicate that concrete reinforcers are more meaningful to the retarded than are more abstract kinds of reinforcers. Another interpretation might be that, although the retarded have suffered a loss in status, the interaction of *Es* with *Ss* may have sufficiently reduced the need for recognition or approval, but did not reduce the concurrent need for material gain. The need for recognition or approval for the nonmaterially rewarded retarded like-



wise was reduced by the interaction of *Es* with *Ss*; and, apparently, the additional nonmaterial reinforcer did little to boost performance above that of the control group. A study in which social reinforcement is eliminated from the experimental situation is needed before the effect of experimentally introduced nonmaterial reinforcers can be assessed properly. The fact that the graduate students who were given nonmaterial rewards reached asymptote at a faster rate than did the groups materially rewarded or not rewarded seems to be related to the past association between performance and letter grade. In addition, the value of the money used would be perceived as far greater by the retarded than by the graduate students. In discussions after the experiment, many of the students stated that as rewards nickels were not sufficient to effect performance change. The lack of differences among the physically handicapped in trials to asymptote may indicate merely that the effects of social reinforcement on drive level (hence on performance) were so great as to override completely the effects of the experimental variables of rewards and incentives.

To test the effect of negative incentives, 30 mentally retarded were retested one month later under all three incentive conditions. Two sequence-of-incentive conditions were used: negative incentives followed by positive incentives or positive incentives followed by negative incentives for the first 10 experimental trials; the mixed-incentive condition was used for both groups for the last five trials. There were no differences due to the type of incentive or the sequence of incentive. The increase in number of trials from 10 to 15 produced no differences among groups.

Epstein (3) suggests to those involved in motivation research that "the usual procedure of working in the laboratory with motivational variables of low intensity and of little significance to the subject, and then attempting to control all sources of incidental variance, is self-defeating, as the latter is impossible" (3, p. 202). He further notes that "when one reads that set effects are more important than drive effects . . . , the conclusion is based on experiments which have worked with a minimal range of drive and a maximal range of set." In this experiment, it appears that neither the instructions nor the rewards were sufficient to override to a significant degree the effects on performance of other sources of drive and of the capacity of the subjects.

### E. SUMMARY

The effects of type of reward, type of incentive, and sequence of incentive on the motor performance of 50 mentally retarded, 43 physically handicapped, and 35 graduate students were investigated. Analyses of variance based on

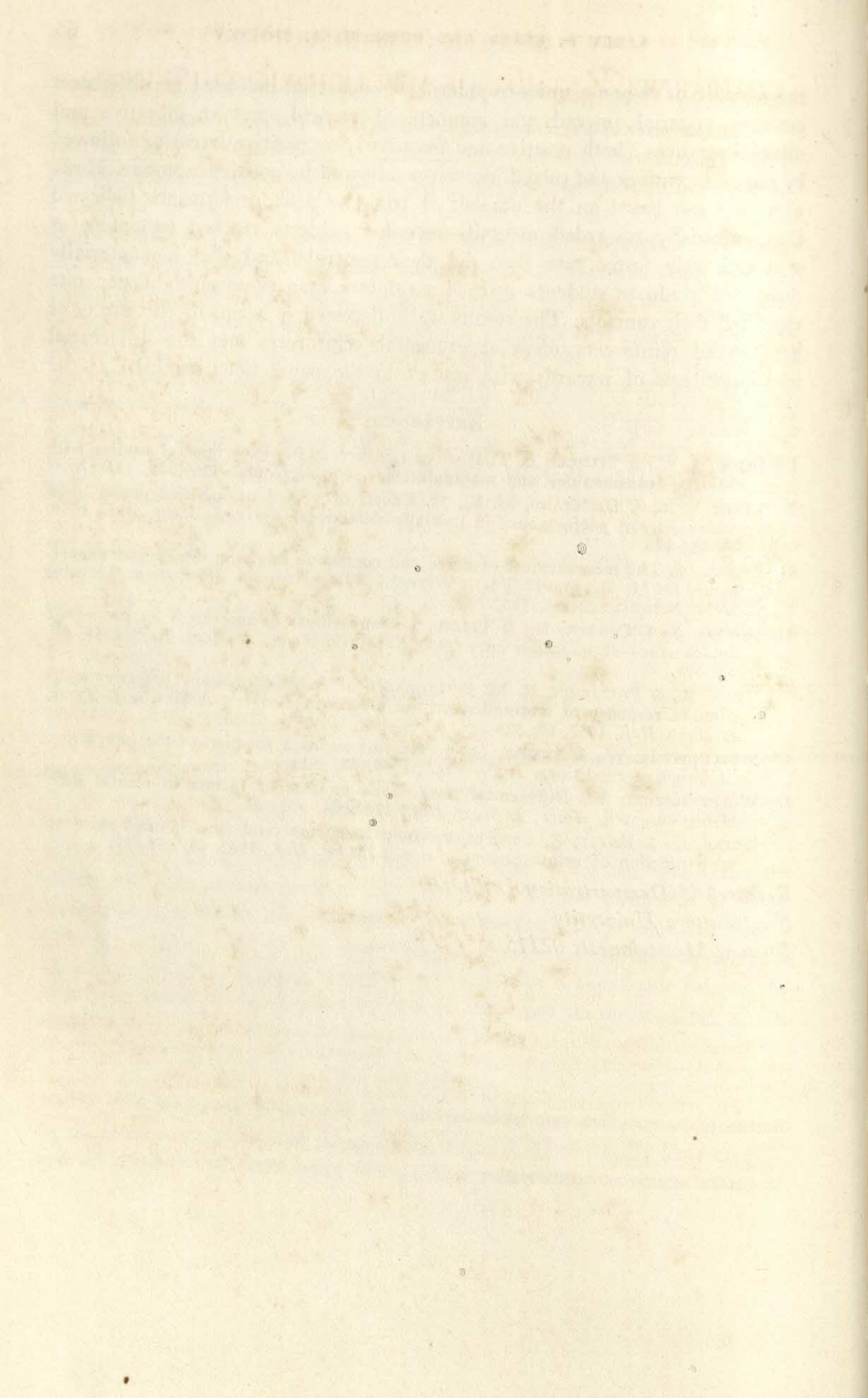
the number of response units completed on each trial indicated no differences between material reward and nonmaterial reward, positive incentive and mixed incentives (both positive and negative), or positive incentive followed by mixed incentives and mixed incentives followed by positive incentive. However, a  $t$  test based on the number of trials to peak performance indicated that materially rewarded mentally retarded subjects reached asymptote at a significantly faster rate than did their controls, and that nonmaterially rewarded graduate students reached asymptote at a significantly faster rate than did their controls. The results were discussed in terms of the effects of generalized reinforcers upon experimental reinforcers and the differential meaningfulness of rewards with respect to the populations studied.

## REFERENCES

1. BIJOU, S. W., & STURGES, P. T. Positive reinforcers for experimental studies with children—consumables and manipulables. *Child. Devel.*, 1959, **30**, 151-170.
2. ELLIS, N. R., & DISTEFANO, M. K., JR. Effects of verbal urging and praise upon rotary pursuit performance in mentally defectives. *Amer. J. Ment. Def.*, 1959, **64**, 486-490.
3. EPSTEIN, S. The measurement of drive and conflict in humans: theory and experiment. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation*. Lincoln: Univ. Nebraska Press, 1962.
4. GORDON, S., O'CONNOR, N., & TIZARD, J. Some effects of incentives on the motor performance of imbeciles on a repetitive task. *Amer. J. Ment. Def.*, 1955, **60**, 371-377.
5. HUNT, B., & PATTERSON, R. M. Performance of familial mentally defective children in response to motivation on the Goodenough Draw-A-Man test. *Amer. J. Ment. Def.*, 1957, **62**, 326-329.
6. STEVENSON, H. W., & SNYDER, L. C. Performance as a function of the interaction of incentive conditions. *J. Personal.*, 1960, **28**, 1-11.
7. WOLFENBERGER, W. Differential rewards as motivating factors in mental deficiency research. *Amer. J. Ment. Def.*, 1960, **64**, 902-906.
8. ZIGLER, E., & UNELL, E. Concept-switching in normal and feeble-minded children as a function of reinforcement. *Amer. J. Ment. Def.*, 1962, **66**, 651-657.

*Research & Demonstration Institute*  
*Northeastern University*  
*Boston, Massachusetts 02115*





## COMPARATIVE STUDY OF METAPHORIC THINKING\*<sup>1</sup>

*Department of Psychiatry, University of Utah College of Medicine*

IJA N. KORNER AND ROGER B. ALLISON, JR.

### A. INTRODUCTION

The concept of body image permits the formulation of interesting deductions. The thinking that centers around body image affords a fruitful approach to the study of certain aspects of the formation of symbolic concepts. Individuals gain body information and learn to utilize their bodies for contacting and dealing with the environment. In the process, parts of the body are vested with meanings and connotations related to their specific functions as agents in the growth and development of one's self-concept.

On the assumption that children are subjected to common developmental experiences, individuals should agree to some extent in their body-image concepts; but although there exists a communality in growth processes, unique and personalized meanings in regard to the body image are also postulated. The second assumption is fundamental for contemporary attempts to interpret figure drawings and is basic to the conceptualization of metaphoric symbolism as that concept is presented in this paper.

The authors consider metaphoric symbolism to be a process of thinking by which an abstract quality, as embodied in an adjective (e.g., good, bad, strong), is translated into an object (e.g., a part of the body). In the present study, the authors attempt to demonstrate that groups of subjects with presumably different backgrounds have differential object responses to a series of stimulus adjectives.

### B. METHOD

#### 1. *Testing Procedures*

Each of the subjects was presented orally with a list of adjectives [see Table 1] and for each word was asked to point to that part of his body that

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 15, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This project was supported by the National Institute of Mental Health (Grant M-1690). Grateful acknowledgment is made to Mrs. Dixie Covey and Mrs. Selma Moskowitz for collecting portions of the data; to Mr. G. Sabourin, Fort Simpson, Canada, for serving as interpreter for the senior author; and to Mr. Paul F. Bramwell for the statistical analyses.



he associated with the stimulus. Subsequent inquiry allowed clarification of the body part and the nature of the particular response. The specific responses (e.g., head, heart, hand) were recorded and classified into one of the categories that follow:

a. *Associative response.* Head, mind, forehead, hair. This category encompasses those responses by which the subject indicates his criterion of choice to be thought or thinking.

b. *Affective response.* Heart, stomach, mouth, eyes, genitals, hips, buttocks, tongue, face, throat, chin, ear, eyebrows, neck. This category encompasses those responses by which the subject indicates his criterion of choice to be primarily emotional or feeling.

c. *Motoric response.* Arms, hands, fingers, shoulders, legs, knees, feet, chest, back. This category encompasses those responses by which the subject indicates his criterion of choice to be movement or action.

d. *Reject responses.* Whole body, side of body, "don't know," goose bumps all over, and refusals. This category may represent a conglomeration of many things; however all responses in this category represent a subject's failure to enter into the task of pointing to that part of his body that he associates with the stimulus word.

## 2. Subjects

Three groups of subjects were tested. Group A consisted of 27 male university students enrolled in a freshman-psychology class; Group B consisted of 20 male university students majoring in accounting; and Group C consisted of 29 male, non-English speaking, nonacculturated, hunting Athabaskan Indians from an isolated region of the Northwestern Territories in Canada.<sup>2</sup>

## 3. Hypotheses and Statistical Technique

The first question is "What effect does a unique culture have upon metaphoric symbolism? Do object responses reflect universal childhood experiences of growth and development? or do they represent manifestations of cultural indoctrination?"

The authors hypothesized that the responses to be obtained from Athabaskan Indians would be significantly different from those of the 27 males in Group A—further, that the more primitive, less differentiated method of thinking by the Indians would produce more reject responses of the don't-know or all-over types.

Of course, the "metaphoric thinking" involved in the study may represent nothing more than the experimenter's cultural background, and the classifi-

<sup>2</sup> See Korner (1) for a detailed discussion of the Athabaskan Indians.

cation of responses may reflect little more than the prevailing customs of Western society. Although intellectual level appears to play only a minor role in metaphoric thinking, differences in educational opportunities and their significance were neither controlled nor studied. Failure to secure an exact Athabascan equivalent for each English adjective also may account for certain differences, but considerable care and cross-checking went into the establishment of the Athabascan Word List.

The authors hypothesized that metaphoric symbolism plays a role in professional choice, perhaps by means of identification and values, and that the modal response category of occupational groups would coincide with certain stereotypes of each occupation. In other words, the authors hypothesized that students majoring in accounting would provide responses significantly different from those of the 27 males in Group A.

The statistical technique used was that of testing with the method of maximum likelihood the groups and response categories for independence of classification. Groups B and C were compared with Group A on each item. At the .05 level of confidence, and with three degrees of freedom, a chi-square value of 7.8 is required for significance.

### C. RESULTS AND DISCUSSION

Table 1 shows the percentages of subjects responding to each category. One of the most obvious conclusions is that individuals differ within as well as between cultures. On each of the 12 stimulus adjectives, Groups A and C

TABLE 1  
PER CENT OF SUBJECTS IN EACH GROUP (A, B, C)\* SELECTING EACH RESPONSE CATEGORY

Stimulus adjective	Associative			Response Affective			Motoric			Reject		
	A	B	C	A	B	C	A	B	C	A	B	C
tired	11	15	10	22	30	14	52	45	7	15	10	69
dangerous	7	35	14	30	25	0	37	25	7	26	15	79
lazy	7	20	49	11	30	3	37	20	3	45	30	45
excited	7	40	24	49	50	21	22	5	0	22	5	55
sad	7	10	45	82	85	10	0	0	0	11	5	45
sick	11	10	14	63	70	7	4	0	0	22	20	79
strong	0	0	0	0	0	3	96	100	34	4	0	63
bad	33	40	73	19	30	10	26	10	0	22	20	17
good	15	45	73	40	40	17	15	0	0	30	15	10
happy	11	15	45	82	85	10	7	0	0	0	0	45
angry	11	35	62	55	45	14	19	20	0	15	10	24
frightened	15	40	21	41	55	65	33	5	0	11	0	14

\* Group A, 27 male freshmen; Group B, 20 male accounting majors; Group C, 29 male Athabascan Indians.



differ significantly at the .01 level; thus the hypothesis that cultural indoctrination has an impact upon metaphoric symbolism is supported.

With one interesting exception (namely, for "bad" and "good") Group C tended to use more responses of the reject category than did Group A.

When Group A is compared with Group B, only three differences (out of 12) exceed the .05 level of confidence. For each of the words "excited" and "frightened," members of Group B gave less variable responses than did the members of Group A. The word "good" is the only stimulus word that produced data in support of our hypothesis that "metaphoric symbolism plays a role in professional choice." In retrospect, it appears that the professional aspirations of the members of Group A may have been too heterogeneous to permit an adequate test of the relationship between metaphoric symbolism and occupational choice.

#### D. SUMMARY

In this study, the authors investigated the process of metaphoric symbolism in three groups of subjects: two groups were reared within a Western culture; a third group, in a remote culture found in the Northwestern Territories of Canada.

Differences between cultures were obtained when subjects pointed to those parts of their bodies that they associated with specified stimulus adjectives. Indian subjects tend in general (three exceptions out of 12 words) to associate words with the whole body or do not know what body part to specify. "Western" subjects provide differential and specific responses. It appears, therefore, that culture has an influence upon the thinking and formation of symbolized concepts.

Our comparison of the metaphoric symbolisms used by freshman students and advanced-accounting students fails to substantiate the hypothesis that body-image correlates of adjectives are related to professional choice. One reason, possibly, for this failure is the fact that a professionally heterogeneous group was compared with a professionally homogeneous group. Comparisons between homogeneous groups might be more productive of positive results.

#### REFERENCE

1. KÖRNER, I. N. Notes of a psychologist fieldworker. *Anthropologica*, 1959, **1**, 1-15.

*Department of Psychiatry*

*University of Utah*

*College of Medicine*

*Salt Lake City, Utah*

## MANIFEST ANXIETY, AMPHETAMINE, AND PERFORMANCE\*<sup>1</sup>

*Department of Student Services, The City College, New York*

---

MARTIN WEITZNER

---

### A. INTRODUCTION

The present experiment had two purposes. The first was to study the effect of manifest anxiety (as measured by the Taylor Manifest Anxiety Scale) on simple and complex tasks. The second purpose was to study the effects of the drug amphetamine in combination with the same variables. Amphetamine increases feelings of mental and physical activation and improves performance on many behavioral tasks (17); therefore the present study posed the question of whether the differences in performance of subjects given amphetamine and those given placebo are analogous to the performance differences of subjects with high- and low-manifest-anxiety scores (12, 13, 14, 15, 16). More specifically, the author sought to determine whether amphetamine improves performance on simple, noncompetitive tasks and impairs performance on complex, competitive tasks.

### B. METHOD

#### 1. Subjects

A group of 420 male undergraduate students were screened with the MAS at Boston University and 96 were selected for the experiment. These subjects were divided equally into a low-manifest-anxiety group (MAS scores 1 to 7), a middle-manifest-anxiety group (MAS scores 12 to 17), and a high-manifest-anxiety group (MAS scores 21 to 36). Each of the three groups was further subdivided so that half of the subjects in each group ( $N =$

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 19, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> Based upon the author's dissertation at Boston University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is indebted to Professor L. J. Reyna, sponsor of the dissertation. The author also wishes to acknowledge the generous aid and guidance of Professors Henry K. Beecher, William G. Cochran, and Gene M. Smith. This work was carried out while the author was a Fellow of the Charles A. King Trust at the Massachusetts General Hospital and was supported in part by the United States Public Health Service, Contract No. USPH-M-987, and in part by the Medical Research and Development Command of the United States Army, Contract No. DA-49-007-MD-76.



16) received amphetamine sulphate (14 mg per 70 kg body weight) two hours prior to the experimental session, and half received visually identical placebo capsules, also taken two hours prior to the experimental session. The drug dosage and time interval between the administration of the amphetamine and the testing was based on implications of earlier work (7, 8, 9, 10, 11).

## 2. Procedure for Paired-Associates Learning

A Hull-type memory drum, Lafayette Model No. 303B, was employed to present a competitive list and a noncompetitive list of paired associates. The noncompetitive list had 15 paired adjectives; the competitive list had 12 (see Table 1). Adjectives for both lists were taken from Haagen's word list (3), and the two lists were constructed to satisfy criteria suggested by Spence (12, 14, 15).

TABLE 1  
STIMULUS AND RESPONSE PAIRS FOR THE NONCOMPETITIVE AND THE COMPETITIVE TESTS

Noncompetitive list		Competitive list	
Stimulus	Response	Stimulus	Response
Agile	Nimble	Ragged	Threadbare*
Gloomy	Dismal	Outworn	Vocal
Empty	Vacant	Tattered	Brave
Tranquil	Quiet		
Mammoth	Oversize	Barren	Fruitless*
Healthy	Wholesome	Desert	Polished
Frigid	Icy	Arid	Jaded
Idle	Lazy		
Joyous	Merry	Fearless	Daring*
Complete	Entire	Valiant	Burly
Distant	Far-off	Gallant	Winged
Pleasant	Genial		
Sacred	Holy	Little	Minute*
Nomad	Roving	Undersized	Ardent
Urgent	Pressing	Petite	Giddy

\* In this table, the stimulus and the response words of the competitive list are arranged so as to clarify the competitive nature of the list. In the experiment, the stimulus and the response pairs for both lists were presented in three random orders.

For the noncompetitive list, each stimulus word was paired with a response word with which it was highly associated; the associative connection between each stimulus word and all other response words was low. Moreover, no stimulus word was synonymous with any other stimulus word, and no response word was synonymous with any other response word.

As shown in Table 1, the competitive list contained four triads. Each triad consisted of three highly associated stimulus words: one response word that was highly associated with the three stimulus words, and two response words that had little or no associative connection with the stimulus words.

Both the competitive and the noncompetitive lists were presented at a two-seconds-on-and-two-seconds-off exposure rate. For both lists, there were three orders of presentation. Half of the subjects learned the competitive list first and half learned the noncompetitive list first.

After the first presentation of the list, the subject's task was to anticipate the correct response to each stimulus word during the two-second interval when the stimulus word only was visible. Subjects were run until they met a criterion of two successive perfect trials or until they had received 20 trials on the noncompetitive list or 35 trials on the competitive list. On each trial, correct anticipations and errors were recorded. An error consisted of no response or an incorrect response during the anticipation interval. Both errors and trials to criterion were analyzed, but because the two approaches yield similar conclusions only the analyses of trials to criterion are reported.

Each subject in the present study (unlike those in earlier investigations) learned both the competitive and the noncompetitive lists; thus it was possible to obtain a measure of what might be called "differential performance": that is, the difference between scores on the competitive and the noncompetitive lists.

### 3. *Procedure for the Simple and the Difficult Digit-Letter Coding Tests*

The simple-coding test consisted of 20 lines of material, such as the following: 4 d P q r M d B 7 W b g J 6 E m N a 5 Z v p L k f h B t 2 V W m. At the top of the page the following code appeared: 2—circle small letters; 3—circle large letters; 4—underline large letters; 5—check on top of small letters; 6—check on top of large letters; 7—underline small letters. The subject's task was to respond to the letters that followed a given number, using the operation for which the number was the symbol.

The difficult-coding test consisted of the same material as the foregoing; however, the instructions were modified to read as follows:

Whenever you see an odd number—1, 3, 5, 7—do *not* respond to the *first* otherwise appropriate letter in each block. Whenever you see an even number—2, 4, 6—do *not* respond to the *last* otherwise appropriate letter in each block. When there is only one appropriate letter in a block do *not* respond to that letter.

For both the simple and the difficult tests, the series of letters following each number was considered one item and contributed one point to the subject's score if all the letters in that item were responded to correctly. Four minutes were allowed for each test. The highest score possible on either was



122 points. Both types of tests were scored on the basis of the number of items right, the number wrong, the number omitted, and the number right minus the number wrong. Only the results of analyses based on the number of items right are presented. As with the paired-associates tests, the fact that each subject took both the simple-coding and the difficult-coding tests made possible a score incorporating both tests: the number of simple items right minus the number of difficult items right.

#### 4. *Procedure for the Pursuit-Rotor Test*

The pursuit-rotor test consisted of a modified phonograph turntable which revolved at 33 rpm. The target was a metal disc one-half inch in diameter inserted in a bakelite disc. The target was three and one-half inches from the center. The performance measure for each subject was the number of seconds on target during each of five 120-second trials.

#### 5. *Order of Test Presentation*

As already mentioned, the 96 subjects consisted of six groups of 16 each. Eight test orders were used. Within each group of 16 subjects, two subjects received each order. This permitted an examination of possible interactions between test order and manifest-anxiety score and between test order and drug effect. (No significant interactions were found.)

### C. RESULTS

#### 1. *Paired-Associates Performance*

Table 2 presents the means and the standard deviations for high-, middle-, and low-manifest-anxiety groups under amphetamine and placebo conditions for the noncompetitive list, the competitive list, and for differential-performance score. The results of an analysis of variance for the performance scores on the noncompetitive paired-associates list are presented in Table 3. The test of the manifest-anxiety-score effect indicates that the three MAS groups have significantly different performance scores. The *t* test for uncorrelated samples was used to evaluate the differences between the group means reported in Table 2. The performance of the high-manifest-anxiety group was found to be significantly superior to that of the low-manifest-anxiety group. Analyses of variance similar to the analysis in Table 3 were also carried out for performance scores on the competitive paired-associates list and for the differential-performance scores: that is, the difference between trials to criterion on the competitive and the noncompetitive list. The test of manifest-

TABLE 2  
 PAIRED-ASSOCIATES PERFORMANCE FOR HIGH-, MIDDLE-, AND LOW-MANIFEST-ANXIETY-  
 SCORE SUBJECTS ON THE NONCOMPETITIVE AND THE COMPETITIVE LISTS  
 (Trials to criterion)

Subjects	List				Differential- performance trials on competitive list minus trials on noncompetitive list	
	Non- competitive Mean	SD	Competitive Mean	SD	Mean	SD
Placebo						
High MAS	7.06	2.11	23.56	5.56	16.50	5.04
Middle MAS	8.56	2.94	21.13	4.08	12.57	4.98
Low MAS	9.56	2.62	19.38	4.85	9.82	6.25
Amphetamine						
High MAS	6.56	2.13	23.31	5.16	16.75	6.69
Middle MAS	7.50	2.15	21.75	3.49	14.25	4.34
Low MAS	7.94	2.64	20.75	4.56	12.81	6.11

TABLE 3  
 DIGIT-LETTER CODING TEST FOR HIGH-, MIDDLE-, AND LOW-MANIFEST-  
 ANXIETY-SCORE SUBJECTS  
 (Number of items right)

Subjects	Test				Differential- performance simple-coding rights minus difficult-coding rights	
	Simple Mean	SD	Difficult Mean	SD	Mean	SD
Placebo						
High MAS	46.06	4.49	22.75	7.38	23.94	7.74
Middle MAS	44.50	9.03	24.38	9.07	20.13	9.16
Low MAS	42.81	8.50	23.13	8.95	19.69	11.52
Amphetamine						
High MAS	49.13	9.65	22.56	8.19	26.56	11.39
Middle MAS	50.06	7.50	23.75	9.29	26.25	8.65
Low MAS	47.63	6.78	25.00	7.20	22.63	7.93

anxiety-score effect on the competitive list yields an  $F$  ratio of 4.25, significant beyond the .05 level. A  $t$  analysis shows the performance of low-manifest-anxiety subjects to be significantly superior to that of high-manifest-anxiety subjects. The analysis of variance for differential performance yields an  $F$  ratio of 7.27 for the manifest-anxiety-score effect, significant beyond the .001 level. The high-manifest-anxiety-score group has a significantly larger differential-performance score than does the middle- or low-manifest-anxiety-score group. This result is a function of the superiority of high-manifest-anxiety-score subjects on the noncompetitive list and their inferiority on the competitive list.



The analysis of variance for the amphetamine effect on the noncompetitive list gives an  $F$  ratio of 3.99, significant at the .05 level (see Table 4). There are no significant differences between the amphetamine and the placebo groups on the competitive list or for differential-performance score.

TABLE 4  
ANALYSIS OF VARIANCE OF NONCOMPETITIVE PAIRED-ASSOCIATES FOR HIGH-, MIDDLE-,  
AND LOW-MANIFEST-ANXIETY-SCORE GROUPS UNDER AMPHETAMINE  
AND PLACEBO CONDITIONS

Source	<i>df</i>	<i>MS</i>	<i>F</i>
Manifest-Anxiety-Score	2	30.80	4.43*
Amphetamine-Placebo (A-P)	1	27.09	3.99*
Order (O)	3	15.18	2.18
Manifest-Anxiety-Score by Drug	2	2.54	.22
Manifest-Anxiety-Score by Order	6	3.99	.57
Drug by Order	3	3.34	.48
Manifest-Anxiety-Score by Drug by Order	6	3.47	.49
Within Error	72	6.96	
Total	95		

\*  $p < .05$ .

There are no significant interactions between the manifest-anxiety-score level and the amphetamine effect (see Table 3). On the noncompetitive list the performance of low- and middle-manifest-anxiety-score subjects is more facilitated by amphetamine than is that of high-manifest-anxiety-score subjects, while on the competitive list the degree of amphetamine facilitation is directly related to manifest-anxiety-score level (see Table 2).

## 2. Digit-Letter Coding Performance

Three additional analyses of variance, also similar to the analysis in Table 3, were carried out on the simple-coding, difficult-coding, and differential-performance scores. There are no significant differences among the manifest-anxiety-score groups on any of the three measures. The amphetamine effect on the simple-coding test gives an  $F$  ratio of 7.88, significant beyond the .01 level. There are no significant differences between the amphetamine and the placebo groups on the difficult-coding test or for the differential-performance score.

## 3. Pursuit Rotor

The theoretical considerations to which reference was made at the beginning of this paper did not yield a prediction as to how the performance of the three manifest-anxiety-score groups might differ on the pursuit rotor. This test was included only for exploratory purposes. In the between-subjects anal-

ysis of variance (see Table 5) an  $F$  ratio of 5.34 was obtained for the amphetamine effect, which is significant beyond the .05 level. None of the other  $F$  ratios reaches an acceptable level of significance. In the within-subject analysis of variance (see Table 6) an  $F$  ratio of 16.30, significant beyond the

TABLE 5  
PURSUIT-ROTOR PERFORMANCE  
(Between-subjects analysis)

Source	df	MS	F
Amphetamine-Placebo (A-P)	1	6630.87	5.34*
Manifest-Anxiety-Score	2	1498.50	1.21
Drug by Manifest-Anxiety-Score	2	217.89	.18
Error	90	1241.85	
Total	95		

\*  $p < .05$ .

TABLE 6  
PURSUIT-ROTOR PERFORMANCE  
(Within-subjects analysis)

Source	df	MS	F
Order (O)	4	48,013.75	2013.16*
Interaction			
Drug by Order	4	388.75	16.30*
Manifest-Anxiety-Score by Order	8	25.25	1.06
Drug by Manifest-Anxiety-Score by Order	8	14.00	.59
Error	360	23.85	
Total	479		

\*  $p < .01$ .

.01 level, was obtained for the interaction of amphetamine and order. The result indicates that with practice on the task the performance of the amphetamine group becomes increasingly superior to that of the placebo group.

#### D. DISCUSSION

The design of the present investigation was such that each subject performed on the noncompetitive and the competitive paired-associates lists; he also performed on the simple-coding and the difficult-coding tests. This method has the advantage of controlling for individual differences in ability and other variables not relevant to the experimental and the theoretical issues under consideration.

The results obtained support earlier findings to the effect that the performance of high-manifest-anxiety-score subjects is superior to that of low-manifest-anxiety-score subjects on noncompetitive paired-associates lists, whereas the low-manifest-anxiety-score subjects are superior on competitive lists.



The major finding of the present study is the limited drive effects of manifest-anxiety level as compared to the effects of the second drive: the drug amphetamine. Amphetamine has statistically significant effects on all three tasks; that is, the drug significantly facilitates performance on the noncompetitive paired-associates list, the simple digit-letter coding test, and the pursuit rotor. These findings are consistent with those in the earlier literature (1, 2, 4, 5, 6, 10, 11). On the other hand, manifest-anxiety level is significantly related only to performance on both paired-associates lists; shows some tendency to influence performance on the digit-letter coding tests; and has no effect on pursuit-rotor performance. The drug amphetamine is more successful than manifest-anxiety level in affecting performance on a wide variety of tasks. Manifest-anxiety level significantly affects performance only when the relative associative strengths of task-relevant and task-irrelevant tendencies are controlled and manipulated (that is, in paired-associates learning). In the present study, then, amphetamine appears to have acted more as a non-specific-drive factor in relation to performance than did manifest-anxiety level.

The author assumes that amphetamine increases drive, thus making it possible to study the relationship of drive to human performance with the direct manipulation of drive in the subject. Direct manipulation of drive has merits not possessed by the indirect method of selecting subjects who are presumed to be different in drive by virtue of difference in manifest-anxiety score.

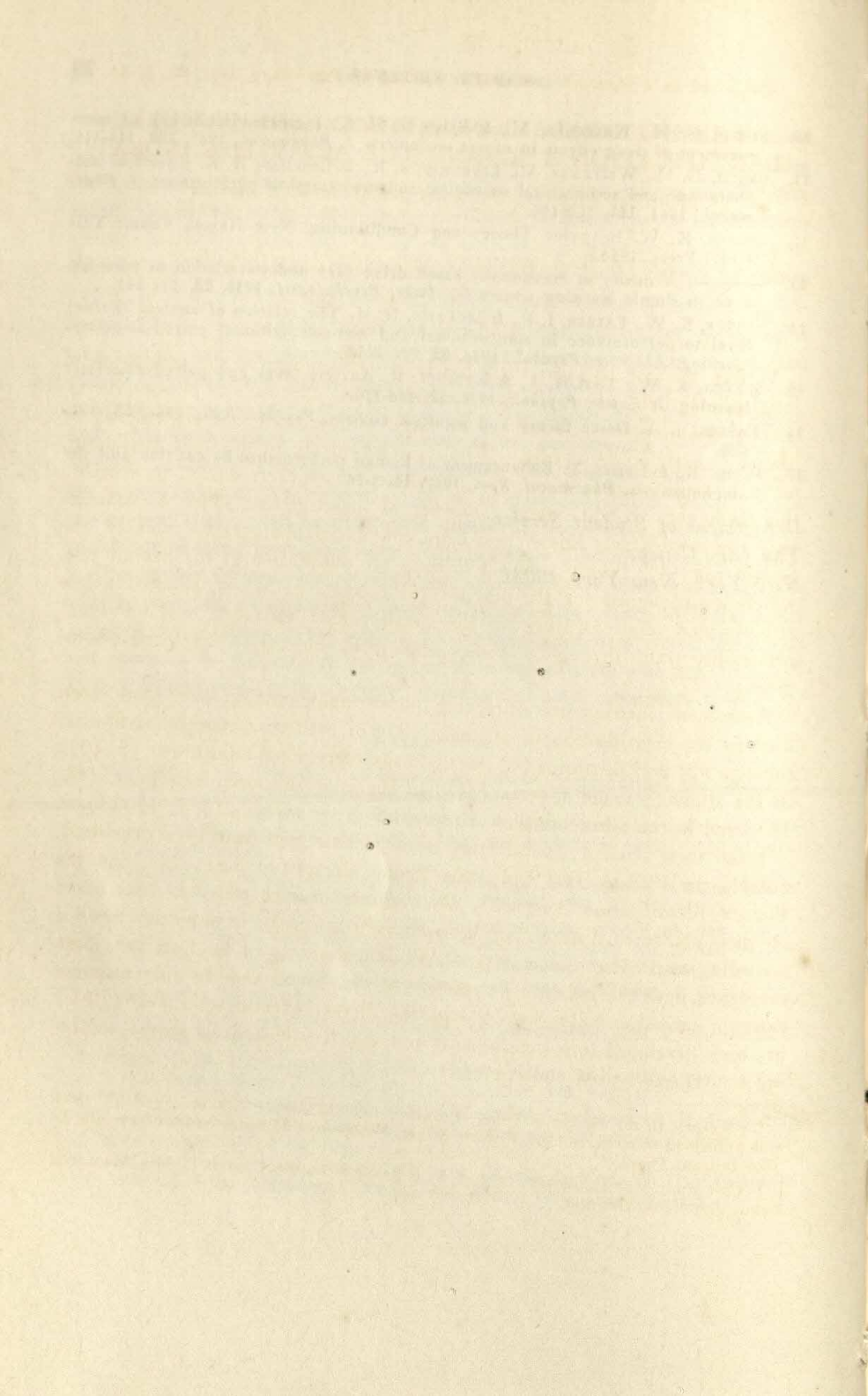
#### REFERENCES

1. BARMACK, J. E. The effect of benzedrine sulfate (benzylmethyl carbrinamine) upon the report of boredom and other factors. *J. of Psychol.*, 1938, **5**, 125-133.
2. EYSENCK, H. J., CASEY, S., & TROUTON, D. S. Drugs and personality: II. The effect of stimulant and depressant drugs on continuous work. *J. Ment. Sci.*, 1959, **103**, 645-649.
3. HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. of Psychol.*, 1949, **27**, 453-463.
4. PAYNE, R. B., & HAUTY, G. T. The pharmacological control of work output during prolonged tasks. Project No. 21-1, 601-0004, Report No. 2, Air University, U.S.A.F. School of Aviation Medicine, Randolph Field, Texas, 1953.
5. ———. Effect of psychological feedback upon work decrement. *J. Exper. Psychol.*, 1955, **50**, 343-351.
6. ———. Factors affecting the endurance of psychomotor skill. *J. Aviat. Med.*, 1955, **26**, 338-339.
7. SMITH, G. M., & BEECHER, H. K. Amphetamine sulfate and athletic performance: I. Objective effects. *J. Amer. Med. Assoc.*, 1959, **170**, 542-557.
8. ———. Amphetamine, secobarbital, and athletic performance. II. Subjective evaluations of performances, mood states, and physical states. *J. Amer. Med. Assoc.*, 1960, **172**, 1502-1514.
9. ———. Amphetamine, secobarbital, and athletic performance: III. Quantitative effects on judgment. *J. Amer. Med. Assoc.*, 1960, **172**, 1623-1629.

10. SMITH, G. M., WEITZNER, M., & BEECHER, H. K. Increased sensitivity of measurement of drug effects in expert swimmers. *J. Pharmacol.*, 1963, **139**, 114-119.
11. SMITH, G. M., WEITZNER, M., LEVENSON, S. R., & BEECHER, H. K. Effects of amphetamine and secobarbital on coding and mathematical performance. *J. Pharmacol.*, 1963, **141**, 100-104.
12. SPENCE, K. W. *Behavior Theory and Conditioning*. New Haven, Conn.: Yale Univ. Press, 1956.
13. ———. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131-141.
14. SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competition and non-competition paired-associates learning. *J. Exper. Psychol.*, 1956, **52**, 296-305.
15. SPENCE, K. W., TAYLOR, J., & KETCHEL, R. Anxiety level and paired-associates learning. *J. Exper. Psychol.*, 1956, **52**, 306-310.
16. TAYLOR, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 301-321.
17. WEISS, B., & LATIES, V. Enhancement of human performance by caffeine and the amphetamines. *Pharmacol. Rev.*, 1962, **14**, 1-36.

*Department of Student Services*  
*The City College*  
*New York, New York 10031*





## THE INTERRELATION BETWEEN THE PICTURE IDENTIFICATION TEST AND THE ACTIVITIES INDEX\*<sup>1</sup>

*Fairleigh Dickinson University and St. John's University*

RALPH B. VACCHIANO AND ROBERT J. ADRIAN

### A. INTRODUCTION

Within the last 10 years there has been a resurgence in the utilization of facial photographs of people as a methodological approach to personality assessment. An impetus to the use of facial photographs was engendered by Murray's early study of fear and its resultant effects on the individual's estimates of maliciousness in others (14). Research workers tended to abandon this technique, though, when a number of investigations failed to substantiate the most widely known test employing facial photographs; namely, the Szondi test (1). Investigations following the publication of the Szondi test, which was designed for the diagnosis of extreme pathology, adopted a less arduous task, concentrating on the assessment of attitudes and individual personality traits through responses of normal subjects to the facial photographs.

Assessment instruments employing photographic stimuli have now been used in the identification of attitudes (8); in trait measurement of normal subjects (2) and neurotics (4); in the study of personal adjustment (5, 10); in the study of value systems (11); in the study of social perception (16, 18); and in the differentiation of psychotics from normals (3, 7, 12, 13).

From these studies, several facial photographic tests have been developed, notably the Picture Test (11), the Photos Preference Test (13), and the Picture Identification Test (6). The validity of such tests has been based on their ability to differentiate between various groups of subjects classified according to predetermined criteria. While a majority of the tests have been concerned with the individual's reaction to the stimuli and the differentiation between normal and abnormal groups, the Picture Identification Test (PIT) has been developed into a personality test which is objectively scored according to Murray's need concept (15).

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 27, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This study was supported, in part, by a grant from the Charles L. Mix Memorial Fund, Americus, Georgia.



Although Chambers has reported success in differentiating a variety of groups with the PIT on some predetermined criteria, relatively little data has been published bearing on its construct validity. The purpose of the present study was to compare the subject's performance on the PIT with a self-report inventory, the Activities Index (AI), since both tests were developed from the same theoretical constructs: namely, Murray's need concept.

## B. METHOD

### 1. *Description of Tests*

The PIT is a personality test intended to measure 21 of the needs formulated by Murray. Test administration has two distinct parts. In the first part the subject is required to look at six sets of facial photographs (either male or female depending on the subject's sex), with six pictures in each set and choose two people he likes best and two he likes least. The same instructions were given for each of the six sets of photographs. In the second part the subject is presented with a list of 21 brief personality descriptions that represent 21 of the Murray needs. The subject must match a person in the photograph for each description he believes best fits that person.

The PIT yields three measures: (a) a judgment score that measures conformity with the norm group in attributing needs to the various photographs, (b) an attitude score that is based on the frequency with which the subject has attributed particular needs to the liked-disliked photographs, and (c) an association score derived from the subject's frequency of associating the same needs with liked-disliked photographs. It is the second measure (the attitude score) that is of particular concern because it is believed that identification by the subject occurs when he attributes the 21 personality statements to the previously chosen liked-disliked photographs. It is assumed that these assignments reflect the subject's own feeling or attitude toward the needs.

The AI consists of 300 items describing commonplace daily activities and feelings that are keyed to the Murray needs. There are 10 items for each of the 30 measured needs, and the subject is required to respond to each item with a "like" or "dislike" choice. An additional 10 needs were derived from the original scale of 30. Validation of the instrument and the scale items was based on factor analysis carried out by Stern (17), the originator of the scale.

Since the PIT and the AI purport to measure the same personality traits and were developed from a similar theoretical rationale (Murray's need con-

structs), it was hypothesized that the subject's scores on the variables measured by the AI would correspond with their attitude scores on the PIT because the same personality traits were measured via different modes of objective evaluation: indirect identification with photographs versus direct self report.

## 2. *Sample and Method*

The sample consisted of 123 male and 168 female college juniors enrolled in introductory-psychology courses. All the subjects were Caucasian, from relatively the same socioeconomic background, and all were maintaining at least the minimum requirements for satisfactory academic standing.

Administration of the PIT and the AI was conducted in two one-hour sessions not more than one week apart. Presentation of the two instruments was randomized with approximately half the subjects receiving the PIT first and the other half the AI. Both the PIT and the AI are machine scored; thus eliminating scorer bias.

## C. RESULTS AND DISCUSSION

Product-moment correlations were computed between the attitude scores of the PIT and the AI for the male and female groups. These correlations were computed for the 20 needs that were common to both instruments. It may be seen from the data presented in Table 1 that for the male group nine of the 20 correlations proved to be significant at the .05 level. These were the correlations for achievement, autonomy, blame avoidance, counteraction, deference, nurturance, order, play, and understanding.

Female subjects proved to be less consistent than males in their responses to the two instruments, with the result that only four of the 20 correlations were significant at the .05 level. These were the correlations for affiliation, sex, understanding, and achievement (negative correlation).

Although nine of the correlation coefficients computed for the males were statistically significant, it should be noted that with the exception of the need "play" for the male group the correlations were not much above the required level needed for significance.

The fact that approximately half of the correlations for males were significant at the .05 level is promising, but the results for females imply that the two tests measure different functions.

Because this investigation was an attempt to determine the construct validity of the PIT relative to its correlation with the AI, it must be admitted



TABLE 1  
CORRELATIONS BETWEEN SCORES ON THE PICTURE IDENTIFICATION TEST  
AND THE ACTIVITIES INDEX

Need variables	Males ( <i>N</i> = 123)	Females ( <i>N</i> = 168)
Abasement	.03	.14
Achievement	.17*	-.17*
Affiliation	.01	.18*
Aggression	-.03	-.06
Autonomy	.17*	.01
Blame avoidance	.18*	-.02
Counteraction	.20*	-.03
Deference	.17*	-.11
Dominance	.08	.05
Exhibition	.10	.13
Harm avoidance	-.12	.11
Inferiority avoidance	.01	.09
Nurturance	.18*	.12
Order	.18*	.08
Play	.35*	-.02
Rejection	-.05	.08
Sentience	.01	-.03
Sex	.13	.15
Succorance	.11	.03
Understanding	.20*	.16*

\* Significant at the .05 level.

that the data do not support the hypothesis that the two instruments are measuring similar dimensions of personality. This is particularly true for females.

For neither test does this study provide information relative to communality with Murray's empirical measurement of needs. It is conceivable that either one or the other instrument independently may have a significant correlation with Murray's need system or that neither instrument does so. It is, however, clear that the two tests do not measure the same variables.

#### D. SUMMARY

The development of the PIT represents a recent trend in the utilization of facial photographs as stimuli in personality testing. Since the PIT is an objective instrument based on Murray's need concept, its construct validity could be directly investigated by employing a self-report inventory, such as the AI, which is also based on Murray's need constructs. The responses of 123 male and 168 female college students were compared on the PIT and AI by product-moment correlation coefficients. Of the 20 measured needs for the male group, nine proved to be significant but low, suggesting little relationship between the two instruments for the male group. Only three

needs correlated significantly for the female group. The modest correlations for both the male and female groups indicate that the PIT and AI are not measuring the same need constructs.

## REFERENCES

1. BORSTELMANN, L. J., & KLOPPER, W. F. The Szondi test: A review and critical evaluation. *Psychol. Bull.*, 1953, **50**, 112-132.
2. CAMPBELL, D. T., & BURWEN, L. S. Trait judgments from photographs as a projective device. *J. Clin. Psychol.*, 1956, **12**, 215-221.
3. CHAMBERS, J. L. Identification with photographs of people. *J. Consult. Psychol.*, 1957, **21**, 232-234.
4. ———. Trait judgment of photographs by neuropsychiatric patients. *J. Clin. Psychol.*, 1957, **13**, 393-396.
5. ———. Trait judgment of photographs and adjustment of college students. *J. Consult. Psychol.*, 1961, **25**, 433-435.
6. ———. Manual for Picture Identification Test. Americus, Georgia: Author, 1964.
7. CHAMBERS, J. L., & BROUSSARD, L. J. Need attitude of normal and paranoid schizophrenic males. *J. Clin. Psychol.*, 1960, **16**, 233-237.
8. DERATH, G., & CARP, F. M. The Picture-choice test as an indirect measure of attitudes. *J. Appl. Psychol.*, 1959, **43**, 14-15.
9. FENSTERHEIM, H., & TRESSELT, M. E. The influence of value systems on the perception of people. *J. Abn. & Soc. Psychol.*, 1953, **48**, 93-98.
10. FIELDS, S. J. Discrimination of facial expression and its relation to personal adjustment. *J. Soc. Psychol.*, 1953, **38**, 63-71.
11. FRYDA, N. H., & CARP, F. M. The Picture Test: Rationale and one validation of the method. *J. Appl. Psychol.*, 1962, **46**, 226-227.
12. IZARD, C. D. Paranoid schizophrenic and normal subjects' perceptions of photographs of human faces. *J. Consult. Psychol.*, 1959, **23**, 119-124.
13. MEER, B., & AMON, A. H. Photos preference test (PPI) as a measure of mental status for hospitalized psychiatric patients. *J. Consult. Psychol.*, 1963, **27**, 283-293.
14. MURRAY, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. Soc. Psychol.*, 1933, **4**, 310-329.
15. ———. Explorations in Personality. New York: Oxford Univ. Press, 1953.
16. SECORD, P., & BEVAN, W. Personality in faces. III. A cross-cultural comparison of impressions of physiognomy and personality in faces. *J. Soc. Psychol.*, 1956, **43**, 283-288.
17. STERN, G. Preliminary Manual: Activities Index—College Characteristics Index. Syracuse, N. Y.: Syracuse University Psychological Research Center, 1958.
18. VANDENBERG, S. G., & MATTESON, E. L. An Experimental Test of Social Perception as Measured by the Interpretation of Facial Expressions; Forms M and F. Schizophrenia and Psychopharmacology Research Project at the University of Michigan and Ypsilanti State Hospital, Box A, Ypsilanti, Michigan, 1960.

Department of Psychology  
Fairleigh Dickinson University  
Florham-Madison Campus  
285 Madison Avenue  
Madison, New Jersey

Department of Psychology  
St. John's University  
Jamaica 32, New York





## EXPERIMENTS ON A POSTURAL AUTOKINETIC EFFECT\*<sup>1</sup>

*Department of Psychology, Tufts University*

---

ZELLA LURIA

---

### A. INTRODUCTION

Normal perception of the position of the body in space is based on both postural and visual cues. Under normal conditions, with an undistorted visual field present, subjects do not err in judging the position of their bodies as stable or moving. The experiments to be reported demonstrate a kind of "postural autokinetic effect": a loss of perceptual accuracy of body position when the visual framework is removed, akin to the usual autokinetic effect that occurs in the absence of visual cues in the dark. The two phenomena share a second characteristic: both are strongly subject to suggestion.

Situations involving contradictory sensory cues give the subject no readily available adjustment to the stimuli, as in the tilted-mirror phenomenon reported by Wertheimer in 1925 [see (10), p. 4]. The contradiction of visual and postural cues often results in the adoption of false but powerfully present visual cues. For large groups of subjects, Witkin (10) has found that an incorrect or misleading visual frame of reference can cause one to lose almost completely his postural frame of reference. By tilting an entire room in which the subject is seated on a chair that can be moved independently of the room, perception of movement (tilting) of the chair is frequently reported, accompanied by reports of the *room* as upright. Such reports may occur despite the absence of any of the pressure or labyrinthine stimulation accompanying body tilt; thus a room tilt of 45 degrees left, with a chair or subject tilt of 45 degrees right, elicited from one subject the report of being tilted 90 degrees to the right in an upright room. Subjects vary from those who seem able to judge body position by postural means and who reject the faulty visual framework to those who accept completely the visual framework. Interestingly, subjects who report dizziness and nausea are those who fall between the two extremes and therefore suffer sensory conflict.

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 27, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> The author is indebted to Drs. H. A. Witkin and W. S. Verplanck for help and guidance in carrying out this study. The pilot work was carried on at Brooklyn College and was supported by a grant from the American Philosophical Society. The final study was supported by the Graduate School of Indiana University.



Piercy (9) has found marked disorientation by subjects asked to match a luminous rod with a standard rod in the dark and in the horizontal plane after a 180-degree body movement. Of interest was the marked discrepancy between subjective confidence reported by the subjects and their objective accuracy.

Many years ago, Hull (6) pointed out the role of reduced cues in augmenting body sway. In his work on sensory conditioning of an auditory hallucination, Ellson (3) noted five favorable conditions for evoking suggestion responses, two of which are relevant in suggestion situations not necessarily making use of sensory conditioning. They are (a) a condition in which the presence or the absence of the objective stimulus is difficult to discriminate—i.e., a condition in which a stimulus is close to its threshold value; (b) a condition in which the subject desires to perceive—i.e., to cooperate with the experimenter or to “do well” in a “perception experiment.” Ellson’s proposed conditions are extensions of the principle that suggestion stimuli can evoke responses when the stimulus conditions are unclear, and the subject is anxious to perceive the experimental stimuli or oblige the experimenter.

Orne’s recent findings in the field of hypnosis (8) lead one to include almost all experimental work in psychology as potentially stimulus ambiguous; thus Orne has been led to suggest more widespread double-blind designs. It becomes of interest, therefore, to study how these effects are transmitted in an experimental situation like that to be reported.

In all of the preceding experiments, as well as in those of the author, the subject is instructed that a certain class of events should be observed and reported. The experimental setup is never clear-cut; i.e., it never allows the subject a means of evaluating the correctness of his response. In other words, the subject’s response is restricted by lack of knowledge; and this lack of knowledge is enforced by the experimenter.

Furthermore, in all situations the experimenter directs the subject’s attention to certain aspects of the experimental setup and limits the possibility of adequate response by giving the subject only isolated items of information. If the experimenter has “planted” bits of information and the subject has almost no information other than that provided by the experimenter, the results are almost certain to be determined (i.e., suggested) by the “planted” information.

In summary, experimental work has emphasized the nature of the stimulus conditions required to call forth responses that are in contradiction with the

objective stimulus situation. First, this may be done by controlling the clarity of the stimuli that the subject perceives. By presenting unclear, ambiguous stimuli we may get responses that are suggested by the experimenter directly, by the apparatus (5), or by the subject's previous attitudes (2). Second, control of the response may be achieved by placing the subject in a limited visual field, allowing contradictory cues to determine the subject's response. How these methods may operate in inducing an illusion of movement is illustrated later.

## B. PURPOSE

The purpose of this research was to determine the efficiency of verbal and apparatus suggestion in eliciting responses to suggestion and to ascertain the nature of responses in the same experimental situation under neutral or non-suggestion conditions.

The experiment also attempted a check of the hypothesis that motion sickness is based in part on the inability of the observer to make adequate use of postural cues in the absence of sufficient visual cues. A subject who reports movement might be expected to be highly susceptible to motion sickness; consequently be expected to score higher than others on a test such as the Wendt Inventory on Motion-Sickness History [see (1)]. Exploratory work on the effect of auditory and cutaneous stimulation on suggestion response was also conducted.

## C. METHOD

### 1. *Apparatus*

The apparatus consisted of an old-model barber's chair, with a built-up wooden back fixed perpendicular to the seat of the chair. The chair had the usual arm rest and footrest but lacked a headrest. The head was free to move, if the subject desired to ignore the instructions. The chair was situated in a light-tight room.

Communication between the subject and experimenter during the period of the experiment proper was provided by a chest microphone to earphones via an amplifying system. This system was unidirectional; the subject could speak to the experimenter, but the experimenter had no means of answering.

A buzzer was used to mask background noise. This was a simple buzzer transformer arranged with a metal cover for the buzzer to amplify the noise. It was placed on a ledge perpendicular to the back of the chair and was directly behind the median plane of the subject's head.



In one of the experiments, a large left shoe with three "electrodes" attached at the instep and at the right and left sides of the shoe was employed. Wires connected the electrodes with a wire leading from the shoe.

## 2. Subjects

All subjects were volunteers from first-year psychology courses. Except in Experiment II, the number of men and women were equal.

a. *Experiments IA and IB.* Two groups of 30 subjects each were run alternately: the first subject served in Experiment IA; the next, in IB; and so on.

b. *Experiment II.* The subjects were 27 students who had served previously in Experiment IB and who had failed to report movement.<sup>2</sup>

c. *Experiments III, IVA, and IVB.* Ten subjects served in each part.

## 3. Procedure

The first experiment was designed to verify the phenomenon of reported movement in total darkness in the absence of true movement after verbal suggestion and to ascertain the effect of neutral instructions on the frequency of reports of movement.

a. *Experiment I (Group A): the experimental group.* The subject was led into a dark room and seated before a long table on which piles of paper, earphones, and an ohmmeter with dials were placed. The subject was seated and told to read the instruction sheet which was as follows:

This is an experiment on the perception of movement. You will be placed in a moveable chair in total darkness. During the period you are in the chair, it will be moved through particular patterns of movement. Observe carefully the directions in which you are moved, for you will be asked to report them. As you notice that you are moved, report your position and state your uncertainty. Report freely. There is no right or wrong answer. We are interested only in how people perceive movement. Report anything else which happens as soon as you notice it. You will be asked questions about what happened at the end of the period.

If the subject asked any questions, the experimenter directed the subject's attention to those lines in the instructions that answered the question. This procedure was followed in all the experiments.

<sup>2</sup> The apparent selection of cases was based on the assumption that these 27 subjects shared a more uniform perception of nonmovement in Experiment IB and therefore could be used to test the question: Can one change the perception of no movement by introducing new suggestion instructions?

Next, the subject was told that the experimenter was going to blindfold him and guide him into the dark room. The subject was led into the experimental room and told:

You are standing beside a chair. This is the arm of the chair [the subject's hand was placed on the arm of the chair] and if you bend down you may feel the seat of the chair. If you lift your foot onto the foot rest you may climb into the chair easily. Good! I am slipping the microphone over your head. Put your feet onto the upper foot rest. [The experimenter moved the subject's legs if the subject hesitated.] Please keep your feet in a relaxed position during the experiment without moving them. That cord you felt just put over your feet is to remind you not to cross your legs. Keep your arms on the arms of the chair. Seat yourself comfortably with your back against the back of the chair and your head free. A buzzer will start in a moment. After that you will hear a door close. This is your signal to lift your blindfold and start reporting into the microphone. Are you ready? Remember when the door closes, you lift your blindfold and you may start reporting.

The buzzer and amplifier were set in operation, and the subject reported for 15 minutes from the dark room. After five minutes, to test whether the subject was selectively or generally attentive, a heavy clamp was dropped; and the subject's verbal report or lack of verbal report of the thud was recorded.

After 15 minutes the subject was again blindfolded, led out of the dark, and told to read the following instructions:

Here are some multiple-choice questions we would like you to answer. The instructions are marked clearly on the top of each questionnaire. Sometimes more than one choice may answer the question. Circle *all* choices, one or *more*, which answer the question. In addition, there may be questions which do *not* apply to you; please write "Does not apply" next to any such questions. Do you have any questions? *Please* do not discuss this experiment with friends or classmates as procedures vary. Your experience may influence the accuracy of their reports. Mr. H. may tell you the results of the experiment when this experiment is over. Again—please do not talk about this experiment.

The subject was then given Questionnaire I, which was answered in the presence of the experimenter.

If the subject answered "yes" to Question 3 ["I felt movement": (a) yes, (b) no], he was given Questionnaire II, III, IV, and V. If he answered "no" to Question 3, Questionnaire II was not given; he was then questioned on light and sound observed during the experiment and on his history of motion sickness. Before completing the questionnaires—to be answered in an office,



not the experimental room—the subject was given a cover-instruction notice reading as follows:

*Remember, answer each question with a choice or choices and/or a note in the margin or "does not apply." Try to make your answers describe accurately the time in the dark. Never go back to change an answer. Write a note, if need be, at the end of the questionnaire.*

Before the experimenter left the subject to answer the questionnaire, the subject was told:

Please read over these instructions. They will remind you of what you read in the other room. Please answer *every* question: if you leave one out, I will think you did not see it and will have to look you up again in class to find out the answer. Also, are you sure you understand why you are not to discuss this experiment with any other student? If you do discuss it, I will have no way of getting the information the experiment is set up for.

The subject was told where to leave the questionnaire and was thanked for his help.

*b. Experiment I (Group B): control group.* The procedure was identical to that described for Group A; however there was a change in the instructions. For Group B, they were as follows:

In this experiment you will spend a certain period of time seated in the total darkness. Report anything which happens as soon as you notice it. You will be asked questions about what happened at the end of the period.

The last two sentences are almost identical to part of the instructions for Group A, but lack the sentences about movement.

*c. Experiment II.* This experiment was designed to determine the effect that instructions suggesting movement would have on subjects who had failed to report movement in Group IB, when no suggestion of movement was given. All subjects in Group IB who had not reported movement were telephoned and told:

The experiment you were a subject for was called an adaptation experiment. Now that I know how you behave in the dark, I wonder if you would be able to be a subject for half an hour in a *perception* experiment.

All subjects agreed, although two did not complete the experiment. The procedure was identical with that for Group IA but for the addition of one line above the instructions given to the members of Group IA. That

added line read, "For subjects previously adapted in the darkroom." Instructions were identical to those used with experimental subjects in Experiment IA.

*d. Experiment III.* This experiment was conducted just as was Experiment IB except that the buzzing sound was eliminated. It was designed as a check on the hypothesis that the buzzing-sound vibration on the back of the chair suggested a motor moving the chair. The procedure in this experiment is comparable with that employed in Experiment IB, with the variable being the presence or absence of the buzzing sound. All instructions relating to the buzzer were deleted.

*e. Group IVA and Group IVB.* The effect of suggestion by apparatus was tested in two groups. Neither group received verbal suggestion. Group IVB was given suggestion "by apparatus." All subjects answered Questionnaires I, II, III, IV, and the Wendt Inventory. Subjects reporting movement in Questionnaire I also answered Questionnaire II. In Group IVA, the procedure of Group IB was followed, with one modification. While the subjects were told to keep their feet "in a relaxed position during the experiment," no cord was put over their feet "to remind you not to cross your legs."

In Group IVB, a shoe was put on the subject. The subject was instructed to remove his left shoe, and his left foot was slipped into a loose shoe with three "electrodes." This shoe was wired to the footrest and was tied with a rubber cord to ensure contact of the foot and the "electrodes." As a means of ensuring that all subjects would relate the shoe and the electric wire, the wire from the "electrodes" was dropped "inadvertently" against the subject's leg before being placed on the floor. The shoe was not discussed with the subject.

#### D. QUESTIONNAIRES

All responses were obtained by six questionnaires as follows: Questionnaire I dealt with general responses of the subject to the darkroom situation, Questionnaire II dealt with the responses of reported movement, Questionnaire IIA dealt with the responses to stimulation of the feet, Questionnaire III dealt with responses to visual stimuli, and Questionnaire IV dealt with responses to sound stimuli. Questionnaire V was the short form of the Wendt Inventory of Motion-Sickness History. To mask the goals of the study, all critical questions were imbedded in many unscored "noisy" questions. In addition, all dependent-variable scores were based on responses to two redundant items.



## E. TREATMENT OF DATA

1. *Scoring Technique*

*a. General.* For all subjects, the response to each question was recorded.

*b. Movement.* The criterion employed for a positive-movement response was positive response to key items: "I felt movement" [(a) yes, (b) no] and "Movement occurred" [(a) never, (b) at rare moments, (c) much of the time, (d) most of the time, and (e) all of the time].

*c. Light.* The criterion for a positive response to light was positive response to any question on Questionnaire III to the item: "I saw a light or lights": [(a) yes, (b) no].

*d. Thud.* The criterion for a positive verbal response to thud was any report of a loud noise at or immediately following the time of the thud. The criterion for a positive questionnaire response to thud was twofold: A response of "yes" to the item, "I heard a sound or sounds other than the stable buzzing sound": [(a) yes, (b) no] plus a choice of (c) on the following question: "At some time I heard the sound (a) of nothing but the buzzer, (b) of a voice, (c) of a loud noise as if something fell, (d) of an elevator rising, (e) of music being played."

*e. Wendt Inventory.* Wendt's scoring technique was used.

2. *Statistical Methods*

All statements on the significance of differences between obtained frequencies are based on the probability values derived from chi square at one degree of freedom.

No statistical treatment of the data on frequency of verbal and questionnaire report of thud is given because the tables adequately show the trends. The Wendt Inventory data are also presented in tabular form for the same reason.

To facilitate the statistical treatment of the results of Group II, a correction has been applied. This correction is based on the assumption that the two subjects in Group IB who did not report movement in Group A, and who were not represented in Group II, would *not* have reported movement in Group II if they had been included. This procedure biases the data against the effect of suggestion by reducing the proportion of subjects reporting movement.

## F. RESULTS AND DISCUSSION

Table 1 demonstrates that the three groups given suggestion instructions or suggestion by apparatus (Groups IA, II, IV) reported the highest fre-

quencies of movement. Even a group that at first—under neutral instructions—gives few responses of movement can be caused to increase their "movement" responses almost four times [see data for Groups IB and II in Table 1]. All differences between suggestion and nonsuggestion groups are statistically significant.<sup>3</sup>

TABLE 1  
FREQUENCY-OF-MOVEMENT RESPONSES

Group	Condition	N	Number of subjects reporting movement	Per cent of subjects reporting movement
IA	Suggestion	30	20	67
IB	Nonsuggestion	30	3	10
II	Suggestion	27	10	37
III	No buzzer	10	2	20
IVA	Control shoe	10	2	20
IVB	Shoe	10	4	40

Only one subject (out of 20) in the group tested for "questionnaire suggestion," by repeating a question at the beginning and at the end of the questionnaire, changed a response as a result of responding to the rest of the questionnaire. In view of the small proportion of changed responses and the use of at least *two* responses as criteria for movement (thud and light responses) we assume we have minimized error by the use of the questionnaire method.

A large error, however, may be introduced by the acceptance of verbal report as a criterion for a particular response rather than some type of questionnaire response. Each group was subjected to the loud sound caused by a heavy object dropped to the floor. Table 2 gives the frequency of spontaneous verbal reports of sound that may be compared with questionnaire reports of the thud. In all groups, the frequency of report on the questionnaire was higher than the frequency of verbal report. This discrepancy between verbal and questionnaire report was largest in Group IA, with 17 of 30 subjects reporting thud in the questionnaire and not verbally. The questionnaire response-verbal response differences in Table 2 lend support to the view that verbal suggestion of the area for response directs the attention of

<sup>3</sup> Chi-square values for differences in frequency of movement were significant beyond the .01 level for comparisons between Group IA and Group IB and between Group IA and Group II; beyond the .02 level for the comparison between Group IB and Group IVB; and beyond the .05 level for comparisons between Group IA and Group III, between Group IA and Group IVA, and between Group IB and Group II. Groups III, IVA, and IVB were run in an exploratory effort. While there was some statistical handling of the data, the results (for the most part) must be considered only suggestive because the samples were small.



TABLE 2  
PER CENT OF VERBAL AND QUESTIONNAIRE REPORT OF THUD  
FOR EACH EXPERIMENTAL GROUP

Group	Condition	N	Verbal	Question- naire	Difference
IA	Suggestion	30	13	70	57
IB	Nonsuggestion	30	73	90	17
II	Suggestion	27	86	96	10
III	No buzzer	10	90	100	10
IVA	Control shoe	10	60	90	30
IVB	Shoe	10	70	90	20

the subject to the kind of stimulus suggested. An extremely low frequency of spontaneous verbal reports of thud is shown by subjects who expected movement (i.e., by those in Group IA). In Group IA, only 13 per cent of the subjects report the thud verbally. The other groups with less explicit instructions range from 60 per cent to 90 per cent reporting the thud verbally. Even on the questionnaire, Group IA remains lowest for the frequency of report of thud. In this group, in which the subjects were "movement oriented," the instructions seemed to limit the likelihood of response to non-movement stimuli; i.e., produced a set against reporting other stimulation. Verbal reports from Group IA were restricted generally to movement or gross sounds. No significant differences between any of the groups was found for the perception of light.

The same extensiveness of verbal report shown by Group IB is seen in another nonsuggestion group (Group III) run under conditions when the buzzer was not sounding. In this group, the verbal responsiveness took the form of detailed reports of auditory stimuli most often exaggerated from unclear, insignificant sounds. Both Groups IB and III show the effects of nonsuggestion instructions in an ambiguous situation; i.e., the subjects give complex responses to relatively simple stimuli.

A comparison of the performance of the subjects in Group IA and that of the subjects in Group IB indicates the efficiency of verbal suggestion in determining responses to an ambiguous situation. The highly significant difference in frequency-of-movement reports between these two groups was 57 per cent, a difference that indicates that movement or suggestion instructions in an ambiguous situation can determine the subject's response.<sup>4</sup>

<sup>4</sup> The pilot study at Brooklyn College gave even more dramatic results. At Brooklyn, the equivalent of our Group IA showed 22 of 30 subjects reporting movement, but 11 of 30 control-group subjects reported movement also. Clearly, sample characteristics may be different.

The results show a consistency of response based on group membership. Although significant differences in frequency-of-movement responses are found between suggestion and nonsuggestion groups (between Group IA and Group IB; Group IA and Group III; Group IA and Group IVA; Group IB and Group II), all those groups run under nonsuggestion conditions (Groups IB, III, and IVA) show no significant differences in frequency-of-movement responses.<sup>5</sup>

Subjects in Group IB who did not report movement were subjects under experimental conditions identical with those for subjects in Experiment II, except for the addition of the suggestion instructions. In Experiment II it is reasonable to expect that Group II will not behave as did Group IA, due to its previous experience in the experiment. Group II consists of "sophisticated" suggestion subjects; the subjects' characteristics change in the direction of Group IA in frequency-of-movement report. This change in the behavior of Group IB when suggestion is introduced appears in the significant difference found in the frequency-of-movement response between Group IB and II, yet the change is not sufficient to eliminate the significant difference between Group IA and II, although both groups are suggestion groups.

While, according to the instructions, both Groups IA and II were attending to movement, the previous experience of the subjects in Group II made them more responsive to the touch than were the subjects in Group IA [see Table 2]. Adaptation to the experimental situation results in consistent differences between the responses of Groups IA and II to questions relating to disturbance. Reports of "head feelings," "stomach feelings," and dizziness occurred more often in Group IA than in Group II. Although no statistical treatment of these data is presented,<sup>6</sup> familiarity with the experiment seems to make for fewer responses indicative of disturbance.

Ten subjects in Group IVA and 10 in Group IVB were run with non-suggestion instructions. Each subject in Group IVB placed his left foot in a loose shoe on which he could feel an electric wire. The results indicate the significance of nonverbal suggestion by the apparatus. Although control sub-

<sup>5</sup> How may this occur? In our procedure the headrest did *not* immobilize the head. Witkin, in a personal communication, suggests that the perception of movement in the dark may be a function of slight head movements, which, when a visual frame-work is absent, can have the effect of serving as cues for body movement. Perhaps, in fact, subjects to whom movement is suggested move their heads more, which in turn produces more cues for body movement. At this point, this is hypothetical, though the latter conjecture is open to experimental check under conditions with head movable and head immobilized.

<sup>6</sup> Since each subject was permitted more than one response, statistical treatment is too cumbersome for the small amount of information to be derived.



jects without shoes and the experimental subjects (Group IVB) showed no significant differences in frequency of responses to questions about the right (nonexperimental or control) foot, significant differences appeared between the groups for frequency of responses on stimulation of the left (experimental) foot. Group IVB shows significantly more reports of special stimulation to the left foot during the experiment than does Group IVA. Less significant differences—producing chi-square values significant only at the .10 level—appear for specific identification of the nature of the “special” stimulation to the foot, by Group IVB.

Some degree of apparatus suggestion is involved in all our experiments. The chair necessarily provided some suggestion because its construction resembled movable chairs, such as barbers’ or dentists’ chairs, presumably familiar to all subjects. But this suggestion is not very powerful, as can be seen by the low frequencies of movement responses in nonsuggestion groups [*see* Table 1 for Groups IB, III, and IVA]. While apparatus suggestion might be expected more frequently among female subjects, on the assumption that they are more “naive” about apparatus, no sex differences of any significance appeared. These results are unusual, for much of the suggestion literature implies that female subjects tend to be more suggestible than male subjects (6).

The hypothesis that subjects responding to the suggestion of movement are likely to be highly susceptible to motion sickness is not verified. On the Wendt Motion-Sickness History Questionnaire, no appreciable difference in the distribution of motion-sickness history scores appear between subjects reporting movement and subjects reporting no movement.

Subjects who reported movement in Experiment IA were almost equally likely to fall in the top half or bottom half of the distribution of Wendt Inventory scores, suggesting the two measures are not related. In the top half, 11 of the 15 subjects reported movement; in the bottom half, nine of the 15 subjects reported movement. No gross sex differences appeared in Wendt Inventory scores or in frequency of reports of movement, a not-surprising result because the largest subclasses for sex consist of only 15 subjects each. It is perhaps notable that one female subject in Group IA reported severe nausea from the rotation and up-down movement she experienced. This evidence hardly warrants a conclusion of sex difference.

As in many studies, under ambiguous-stimulus conditions subjects attempt to structure the situation for themselves and each subject does this in unique fashion; thus opportunity is provided for an expression of the role of per-

sonal factors. As an extension of this point, if the stimulus situation is ambiguous, a statement from the experimenter (assuming, as one may, that the subject accepts the truthfulness of the experimenter's statement) may be used by some subjects as an aid in structuring the situation. If his statement has effect we assume suggestion, but the phenomenon may not be as mechanical as the word "suggestion" implies and it may have a reasonable adaptive basis. In effect, so-called suggestions may define the nature of the category to be judged under circumstances such that the category is too big to permit a single value judgment. In Asch and Lewis's work on judgment of persons (2), shifting the concept of the occupational category lurked behind suggestion effects. In our study, the experimenter's instructions change the category from perception in general to perception of movement in particular.

We have been studying perception of body position in the absence of the visual field ordinarily available. We have done this by eliminating the various reference points that usually serve to "fix" body position, making it harder for the subject (or at least for some subjects) to determine body position. We know that in other circumstances perception in the absence of a visual field is unstable and effects, like head movement and eye rolling (which, with a full field, matter not at all), become influential. The experimenter's statements give the subject aid in judging *what* is happening under difficult conditions of judgment.

At another level, the results of this study indicate that suggestion effects can be very powerful. Experimenters can help focus subjects' responses by direct instruction and by apparatus. Orne's results (8) lead one to suppose that "naive" subjects have fairly definite conceptions of the purposes of psychologists in conducting the experiments. The corrections of the subject's conception is not the issue—the impact of the subject's conception of the purpose of an experiment on his responses is the critical concern.

Frank's recent discussion of psychotherapy in the framework of older conceptions of suggestion, witchcraft, and influencing techniques (4) raises many interesting questions. Frank believes that suggestion effects are built into psychotherapy, and that the therapist's authority plus his implicit promise to try to help his patient convey to susceptible patients the clear suggestion that they will be made to feel better. Subjects who respond to placebo effects respond to just such suggestions.

It may be that suggestion affects much wider areas of human behavior than we have thought heretofore. McGuigan's report (7) stressing the need to control "experimenter effects" is another addition to the growing body of



data indicating that human variables on the stimulus side may help direct subjects' responses more often than we like to believe.

### G. SUMMARY

Six groups of subjects responded to a darkroom situation with and without suggestion of movement introduced by the experimenter. One group, given verbal suggestion of movement, reported movement significantly more frequently than did groups which received no verbal suggestion of movement. A second group, given neutral instructions, reported significantly fewer movement responses than did all groups which received suggestion instructions. A retreat of subjects in the second group, with the suggestion of movement, showed that both the suggestion and the previous experience affected the group's movement responses. A third group yielded results indicating that the use of auditory stimulation did not determine the frequency of movement response. The final groups present evidence of the efficacy of apparatus as a means of introducing suggestion of response. Subjects given suggestion instructions were found to be less responsive to stimuli unrelated to the suggestion than were other experimental groups; suggestion instructions seem to define the limits of the subjects' attention. Subjects reporting movement were found to be no more highly susceptible to motion sickness than were subjects not reporting movement.

Results are discussed in terms of the use made by the subject of instructions to help structure ambiguous-stimulus conditions. A possible role played by small head movements in perception of body movement is also suggested.

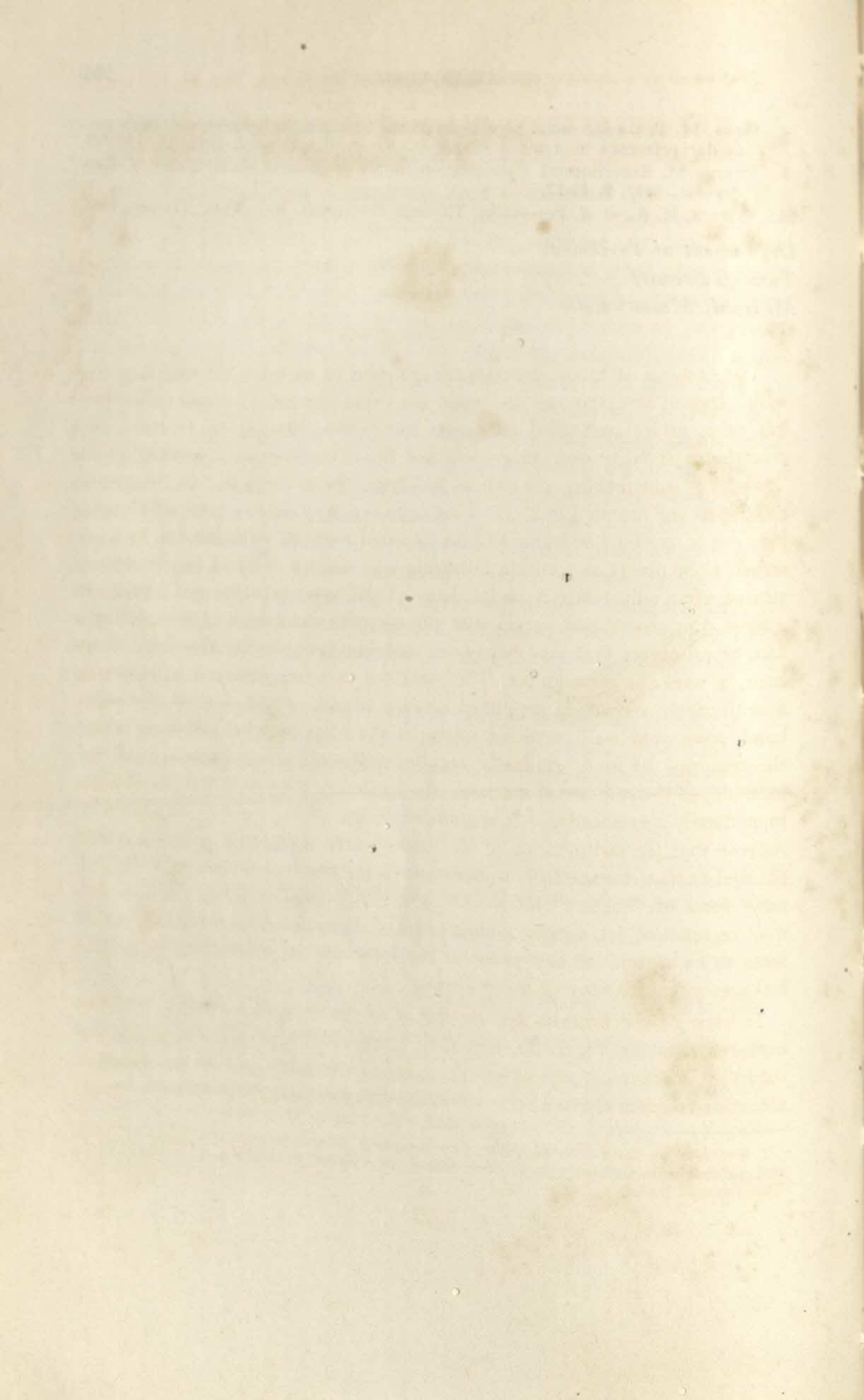
### REFERENCES

1. ALEXANDER, S. J., COYNE, M., HILL, C. J., JR., RICCIUTI, N. A., & WENDT, G. R. Wesleyan University Studies of Motion Sickness: VI. Prediction of sickness on a vertical accelerator by means of a motion sickness history questionnaire. *J. of Psychol.*, 1945, 20, 25-30.
2. ASCH, S. E. Studies in the principles of judgments and attitudes: II. Determination of judgments by group and by ego standards. *J. Soc. Psychol.*, 1940, 12, 431-445.
3. ELLSON, D. G. Hallucinations produced by sensory conditioning. *J. Exper. Psychol.*, 1941, 28, 1-20.
4. FRANK, J. D. Persuasion and healing: A comparative study of psychotherapy. Baltimore: Johns Hopkins Univ. Press, 1961.
5. HENON, W. T. The group demonstration of illusory warmth as illustrative of the phenomenon of suggestion. *J. Abn. & Soc. Psychol.*, 1927, 27, 341-344.
6. HILL, C. L. Hypnosis and suggestibility: An Experimental Approach. New York: Appleton, 1911.
7. MCGUGAN, F. S. The experimenter: A neglected stimulus object. *Psychol. Bull.*, 1945, 60, 421-428.

8. ORNE, M. T. On the social psychology of the psychological experiment: with particular reference to demand character. *Amer. Psychologist*, 1962, 17, 774-781.
9. PIERCE, M. Experimental disorientation in the horizontal plane. *Quart. J. Exp. Psychol.*, 1957, 9, 65-77.
10. WITKIN, H. A., et al. *Personality Through Perception*. New York: Harper, 1954.

*Department of Psychology*  
*Tufts University*  
*Medford, Massachusetts*





## THE RETENTION OF SENSORY EXPERIENCE: IV. SHORT-DELAY VERSUS LONG-DELAY INTERVALS\*

*University of Pittsburgh School of Medicine*

---

H. E. KING

---

### A. INTRODUCTION

The findings of recent experiments designed to measure the accuracy with which human observers can find again the value of sensory stimuli experienced but once, under controlled laboratory conditions, have given evidence of a remarkable stability over time—whether the delay between experiencing the “standard” and making a match to its “trace” be as brief as two minutes or as long as one month (1, 2, 3). That is to say, a successive comparison made between a *standard* stimulus and an adjusted *variable* stimulus has been observed to be nearly as accurate following a prolonged delay as can be demonstrated after only relatively brief delay at the same experimental sitting. In a sense, this observation agrees with the common experience of recognizing a face or an object that may have been seen but once despite the delay of an hour, a week, or even longer. The sustained accuracy observed in matching quantitatively controlled simplified sensory stimuli would not, on the other hand, have been well predicted solely on the basis of what is known about the retaining of more gradually acquired subject matter—as is typical, for example, of learned verbal material. Since primary sensory experience is more immediately apprehended (5, pp. 38-39; 9, pp. 393), there is no reason to suppose that the continuation of its “traces” need necessarily follow a course parallel to that characteristic of more gradually acquired information. Indeed, some form of “holding” mechanism whereby immediate sensory impressions may be retained for a time without marked distortion by assimilation would seem to be a requisite first phase in the formation of more complex pattern linkages or associations of many kinds.

In view of the firmness and regularity of the delayed matching behaviors explored thus far (1, 2, 3), and their unexpected accuracy and persistence over time, the present experiments have been undertaken to turn experimental attention to much shorter delay intervals with two specific purposes in mind:

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 27, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



(a) to observe the point of origin in time of those constant errors (or systematic distortions of subjective equality) that have been observed with greater delay and (b) to replicate the earlier findings in part, using independent subject groups working at the same tasks; while placing an emphasis on the influence that may be exerted by elements of the test methods themselves on the form of the data recorded.

## B. METHOD

The modalities chosen for study, visual and auditory, and the stimulus dimensions singled out for special observation within each sense (intensity, frequency, and duration) are the same as those applied earlier to delays of much longer duration. Each of the six test methods is summarized below; a detailed description of apparatus and procedure has been described elsewhere (1, 2, 3).

### 1. *Visual Intensity (Brightness)*

Each subject, after 10 minutes of adaptation to the low, ambient illumination of the darkroom in which he was seated, was asked to look at a round test patch on a panel before him, using both eyes. He was told that after a "ready" signal the patch before him would be lighted for five seconds, and that he was to notice "how bright it was." After the stimulus went off and after the passage of a predetermined delay interval, the subject was asked to approximate the brightness of the foregoing *standard* stimulus by adjusting a rotary, manual control that continuously varied the intensity of a second *variable* stimulus lighting the same test patch. The standard brightness was 74.5 mL, and the visual angle of the test patch  $3.3^\circ$ . Range of adjustment of the variable field was five to 179 mL.

### 2. *Auditory Intensity (Loudness)*

A similar disposition, with the subject seated in a sound-deadened chamber and asked to attend to the loudness of a standard tone (600 cps, 43.8 db), sounded for five seconds. After a predetermined delay, the subject was asked to duplicate the loudness of the original tone by adjusting a manual control that would continuously vary the loudness of the comparison tone within a range of zero to 100 db.

### 3. *Visual Frequency (Flash Rate)*

Following adaptation to dim ambient light, the subject was asked to inspect a test patch that, after a ready signal, was to be illuminated by a flick-

ering light. He was to form an impression of this fixed rate of flash (25.0 cps) during a five-second inspection period and, after specified delay, to try to reproduce this rate of flash by the adjustment of rotary frequency controls. The adjusted flash rate varied between the lowest delivered by the instrument and the individual critical fusion frequency of the subject for this stimulus; usually five to 50 cps.

#### 4. Auditory Frequency (Pitch)

After a five-second exposure to a standard tone (1000 cps, 75.0 db), with instruction to notice its characteristic "note" or pitch, the subject was asked to try to reproduce this tone, following a delay, from among a continuously variable range of matching tones in the audible frequency range of 20 to 20,000 cps.

#### 5. Visual Duration (Flash Length)

Each subject was exposed to the dim flash of a test patch on a panel before him, following a ready signal, and asked to notice "how long it remained on." After a predetermined delay, he was shown how he might relight the same test patch by closing a telegraph key and was asked to attempt to match the two-second duration of the prior standard.

#### 6. Auditory Duration (Tone Length)

Much the same as Test Method No. 5, but the interval was delimited by a standard tone of 1.5-seconds duration (200 cps, 80 db). Following a specified delay the subject, by pressing a reaction key, attempted to reproduce this exact interval by sounding the tone again for an interval of his own choosing. Again, the response could vary—at least in theory—between zero and infinity.

### C. SUBJECTS AND PROCEDURE

The experimental variable for all six experiments is the interval of delay between the discontinuation of the *standard* and the beginning of the matching adjustment of the *variable* stimulus by the subject. Three relatively short intervals were chosen for study: 15, 30, and 60 seconds. These intervals provide samples of sensory-matching behavior for periods one-half that of the shortest delay used in prior experiments (two minutes), one-half of that (30 seconds) and half that again (15 seconds).

Forty-eight subjects, student nurses, were observed. Subjects in each of three groups ( $N = 16$  in each group) adjusted the variable stimulus to match its "memory" of a standard following a delay of 15, 30, or 60 seconds. Ran-



dom sequences were followed in determining the factors of test order, delay interval, and the stimulus value from which the resetting of the variable stimulus was begun. To avoid possible complication by a confusion of "traces," no subject made more than one match in a given stimulus dimension nor after more than one interval of delay; therefore the results consist of group findings at each delay interval, with wholly different subjects included in the groups for each period of delay. Student-selection procedures provide a reasonable balance in group composition in terms of age, sex, and educational experience. As subjects were drawn from the same source and had to meet the same criteria as those used in earlier, related experiments (1, 2, 3), they may be expected to yield approximately similar behaviors. All groups were 100 per cent samples: i.e., complete first-year nursing classes.

#### D. RESULTS

##### 1. *Intensity (Brightness and Loudness)*

The intensity of the standard stimuli and the mean intensity for all resets of the variable stimuli are given in Table 1 and have been grouped for visual inspection in Figure 1. Just to the right of the plot of these new data

TABLE 1  
STANDARD STIMULUS VALUES AND MEAN RESET PERFORMANCE AFTER DELAY

Stimulus dimension	Standard stimulus	Variable stimulus		
		15 seconds	30 seconds	60 seconds
Intensity				
Brightness (in mL)	74.5	70.5	87.6	69.6
Loudness (in db)	43.8	42.8	44.7	46.3
Frequency				
Flash rate (in cps)	25.0	24.1	26.9	29.2
Pitch (in cps)	1000	1024	1286	1244
Duration				
Flash length (in secs)	2.00	1.53	1.47	1.64
Tone length (in secs)	1.50	1.19	0.98	1.01

and drawn to the same scale are the values previously obtained with identical test methods and standard stimuli, following delays of two minutes and one to 28 days (1).

All of the matches made to a standard brightness show a close approximation to its known value. The full range of response possible (five to 170 mL) is not shown on the ordinate of Figure 1(A), and it is to be noted that the mean reset values fall within 15 mL of the correct (standard) value after all intervals of delay. There is a slight reduction in mean adjusted value at

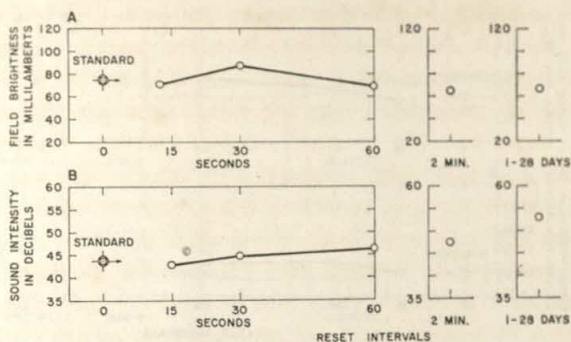


FIGURE 1

BRIGHTNESS AND LOUDNESS OF THE STANDARD STIMULI AND MEAN  
RESETTINGS OF THE VARIABLE STIMULI AFTER DELAY

[The line-connected open circles at the left represent the data of this experiment; the single open circles at the right provide comparison data obtained with longer delay intervals (1).]

15 seconds, followed by an increase at 30 seconds, and another slight reduction after 60 seconds of delay. None of these variations differs significantly (by *t* test) from the standard value nor do they differ reliably from one another. Both the central values and the range of variation made by these subject groups, after delay intervals of less than one minute, are in keeping with the close approximation of the standard found after longer (two-minute) and very much longer (one-day to 28-day) periods of delay (1).

Matches made to the loudness standard, shown in Table 1 and Figure 1(B), fall close to the standard after a 15-second delay; then were found to be increasingly higher following delays of 30 and 60 seconds. None of these values differs significantly from the standard nor do they differ reliably from one another. Only the middle range of possible resettings (zero to 100 db) is shown on the ordinate of Figure 1(B), and it is to be noted that all reset values fall within three db of the correct (standard) value. The rise observed in mean loudness adjustment with increasing delay appeared to be consistent with the finding of significantly raised values previously reported (1) and reproduced in summary at the right of Figure 1(B).

## 2. Frequency (Flash Rate and Pitch)

The flash rate of the standard stimulus and the mean frequency of the resettings made of the variable stimulus following delay are shown in Table 1 and have been graphed in Figure 2 (A). The adjustments made of flash rate after 15 seconds of delay fall slightly below, but not significantly below, the



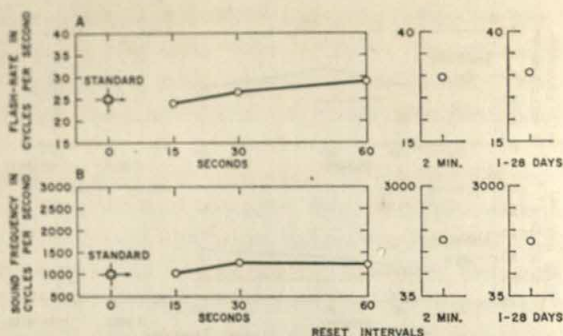


FIGURE 2

FLASH RATE AND PITCH OF THE STANDARD STIMULI AND MEAN  
RESETTINGS OF THE VARIABLE STIMULI AFTER DELAY

[The line-connected open circles at the left are data from this experiment; the single circles at the right provide comparison data obtained with longer delay intervals (2).]

standard; while those made after 30 seconds and 60 seconds of delay rise above the standard rate (significantly at the 60-second interval,  $p < .01$ ). Although neither the mean resettings of the variable stimulus at 30 seconds nor at 60 seconds falls significantly above the value of the preceding interval, a trend toward a rise in average frequency with increased delay seems apparent; and the rise to a value significantly above standard by the 60-second delay interval is consistent with the significant increases previously observed after delays of two minutes and one to 28 days (2).

The matches made to an auditory standard frequency, shown in Table 1 and graphed in Figure 2(B), also begin at a level approximating the standard value after 15 seconds of delay; then were found to rise slightly for the groups tested at 30-second and 60-second intervals. These increases do not differ reliably from the standard value nor from one another. The upward trend does conform to the earlier reported significant rises in average resettings after delays of two minutes and one to 28 days (2). All adjustments made by the subject groups of this experiment fall within 300 cycles of the standard, within a possible range of 20 to 20,000 cps.

### 3. Duration (Flash Length and Tone Length)

A consistent reduction in the reproduction time of both flash-delimited and tone-delimited standard durations may be noted after all intervals of delay in the mean values given in Table 1 and graphed in Figure 3 (A and B). Reproductions of flash length fall significantly below the standard for the sub-

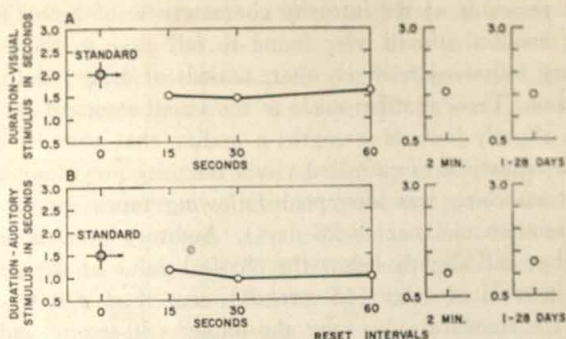


FIGURE 3

FLASH LENGTH AND TONE LENGTH OF THE STANDARD STIMULI AND MEAN RESETTINGS OF THE VARIABLE STIMULI AFTER DELAY

[The line-connected open circles at the left are data from this experiment; the single open circles at the right provide comparison data obtained with longer delay intervals (3).]

ject groups tested after 15-second, 30-second, and 60-second delay intervals ( $p < .01$ ); no one of these groups differing reliably from another. Reproductions of tone length, although lower after 15 seconds of delay, are not significantly below the standard duration; while reproductions made after 30-second and 60-second delay periods do achieve significance ( $p < .01$ ), but do not differ reliably among each other ( $t$  test). These consistent reductions in the reproduction time of both visual and auditory standard-stimulus intervals are in accord with the significant reductions observed for these stimuli after longer delay intervals (3).

### E. DISCUSSION

An impression of stability in function emerges from an overall inspection of these data, expressed by the accuracy with which variable stimuli have been adjusted to approximate the "traces" of sensory standards experienced earlier—whether these matchings were made after either very short or longterm intervals of delay. When errors have appeared among the resettings, they have been found to be systematic in both direction and amount—and these deviations will be examined in detail for what they may suggest about the influence of delay on retained sensory impression. A focus on the kinds of error observed ought not to distract attention, however, from the fundamental demonstration of the ability (accuracy) shown by the average subject in finding close approximations to a variety of sensory values he has experienced but once briefly under controlled laboratory conditions.



The estimates made of the intensity characteristic of both the visual and the auditory standard stimuli were found to fall close to the physical value of the standard following relatively short periods of delay: of 15, 30, and 60 seconds duration. Those matches made to the visual standard stimulus tended to be of only slightly lowered strength; a finding that accords well with the nonsignificant reductions of estimated visual intensity previously observed (1) when similar matching was attempted following much longer intervals of delay (two minutes and one to 28 days). Auditory-intensity matching, although found to fall slightly below the physical value of the standard after the shortest interval of delay (15 seconds), was then seen to rise progressively above the standard value over the longer (30-second and 60-second) delay intervals. This trend toward a gradual increase in estimated auditory intensity over short-term delay is clearly consistent with the observation of significantly raised intensity matches to auditory stimuli that previously has been described when estimates were made after longer periods of delay (1). The difference in direction (sign) of the constant error is a generally consistent one for the two modalities and appears to reflect a basic difference in the stability with which rudimentary impressions of visual and auditory intensity may be stored. The more systematic rise (and ultimately more significant rise) in the average value of auditory-intensity matching presents a rather more striking pattern than does the only slightly reduced, nonsignificant and less progressive lowering of average visual-intensity matching through time. An analysis of matching performance in terms of the geometric mean, undertaken to permit a more appropriate evaluation if the error found tended to be constant with respect to stimulus magnitude (8), does not change the trends just described: The central tendencies of the matches so calculated shift in fact only slightly (but consistently) in the direction of lessened error: i.e., to a slightly more accurate approximation of the intensity of the standard stimuli in both modalities.

The delayed successive-comparison matches made to the frequency characteristic of visual and auditory standard stimuli were also observed to fall close to standard values after the shortest delay period (15 seconds) and then to rise gradually over the longer intervals of 30-second and 60-second duration. This trend toward a rising matched frequency over time reached a level significantly above the standard only for the visual sense at the 60-second interval. The regular trend toward matching at a higher frequency, although not usually reliably increased for either visual or auditory senses, does conform well with the significant increases reported earlier among frequency

matches made to standard stimuli in both modalities following longer intervals of delay (2). As had been found for the intensity characteristic, a consideration of the frequency-matching data in terms of the geometric mean produced no important departure from the patterns described by the arithmetic average.

The reproductions of standard stimulus durations, delimited either by visual or auditory signals, present an unmistakably similar pattern for the two modalities observed—but a pattern quite different from that just described for intensity or frequency matching. The key to this difference appears to rest in the unusual starting-position effect of temporal reproduction (3, p. 304). A tendency for judgments of equivalence to be displaced in the direction from which search is begun is commonly found in sensory experimentation (5, p. 44). This tendency toward “early arrest” is particularly clear in judgments made by the method of successive comparison, but is usually neutralized by alternating the point from which search is begun—with beginning values first chosen from “above,” then “below,” the true equivalent. As time moves only forward, all approaches to estimating a standard stimulus of *duration* by reproducing it necessarily move forward. That is to say, all temporal reproductions begin at the common point of zero time; therefore, any tendency toward early arrest would produce stopping short of the true value: an underestimation. All matched equivalents of duration averaged for this series were in fact less long than the standards delimited either by visual or auditory stimuli. The degree of foreshortening was much the same for brief delay as had been observed earlier when similar estimates were made over longer periods of delay. No clear trend emerged toward a systematic change in temporal estimation with respect to the delay between the standard and reproduced (variable) stimuli. Both the direction and the amount of the constant error of temporal matching at all delay intervals was clearly evident in the data samples collected within 15 seconds following the presentation of the original, standard stimuli.

The main finding of the six experiments performed (to summarize the common trend among tests of different dimensions in differing modalities) appears to be the unusual accuracy and general stability of the matches made by variable stimuli to foregoing standards—despite short or relatively great delay between experiencing the original and the judging of its match. The resettings often fell close enough to the original physical values to suggest more nearly the phenomena and problems met in establishing a difference limen (where delay is virtually nonexistent) than they seem to reflect a systemati-



cally reduced ability through time or the uncertainty or chaos in reproduction that might have been anticipated. There even appears to be a tendency toward a growing positive constant error with time of delay, but the evidence for the latter generalization is less clear. The most obvious exception, seen among the data on temporal duration, has been thought to reflect a unique influence of procedure on result (a one-ended starting-position effect) and not to pose a clear exception. Each of two broad generalizations, based on all of the available data [(a) obtrusive stability through time for most measures of performance and (b) the positive-tending nature of the relatively minor constant errors observed with delay], merits brief comment on its possible meaning and implication for future research with retained sensory experience.

The terminology of retaining has become so firmly identified with the study of learned material—typically gradually acquired and gradual in its loss—that one may come to expect a similar pattern to govern other forms of information acquired or held by the living organism. The idea of an energy exchange implied by the words "living organism" may tend to suggest some necessary form of energy loss through time that must unfailingly result in lessened "traces" with the passage of time. It does not follow, however, that either of these persuasions needs to apply to the reception and (at least) temporary holding of sensory impressions. Being more or less instantly apprehended (5, pp. 38-39; 9, pp. 393) and placed with great accuracy in reference to the biological limits of sensitivity of the receiver (7, 8), they may be "held" or resupplied with the energy needed for their continuation (just as an insight may be held or a goal, or an imprinting). Neurophysiological investigation has provided examples of such action or its analogues in the observations made upon the noncorrespondence between sensory stimuli in the environment and its neural consequences monitored at various points along the course from reception at the end organ to cerebral projection (6). An interpretation of these disparities has been offered in terms of "active neural mechanisms resisting change," and hope has been expressed that an "analysis of the effects of prolonged stimulation may provide experimental access to some of the mechanisms underlying memory, learning and other phenomena relating to sensory experience" (6, pp. 430). While the findings of the present experiments on organized human sensory retaining may be related only in the most general way to the available neurophysiological data, something of the same functional end effect of resistance to change appears to typify the samples of rudimentary sensory-experience retaining under investigation here. Whatever the mechanisms, the general adequacy of delayed sensory matches

recorded have proved to be close enough to original stimulus physical values to lend empirical weight to the view that sensory reception and its direct storage is governed by its own time constants—not by those applying to the retaining of more typically “learned” material.

The tendency toward minor positive error (augmented value) seems particularly puzzling as it may be thought to express an even greater resistance to loss through time for this type of material while held in storage. A re-examination of all data by computing the geometric means of matching performance—to permit a better inspection of error in ratio to stimulus magnitude—does not reduce this tendency appreciably. The slow but regular growth in constant error through time in more than one modality, using identical test procedures with different subject groups, suggests its possible generality and makes unlikely an explanation by a purely procedural artifact. As it has been found that delay does not seriously affect comparison, the sort of procedural factors that may exert an influence on the establishment of difference limens, naturally, will demand much closer attention in future research. For example, the subject's search for matched equality of the variable with the standard stimulus, beginning at values clearly discrepant from it, may produce early arrest as soon as the zone of nondetectable difference (less than a jnd) is entered. Since the absolute value of the indifference zone presumably increases as a function of stimulus magnitude, the zone of nondetectable difference would be expected to be greater for approaches begun at values higher than the standard than for those started below it. The refinement of procedure by psychophysical variation where great delay is not the experimental variable may be depended upon to clear these possible sources of confusion in future experimentation: e.g., by relating the jnd's established by stimulus increment to the starting-position errors of the method of successive comparison. The stimuli used in the present experiments fall in the “comfortable range” of apprehensible stimuli—requiring neither strain to detect nor tolerance of pain resulting from stimulating intensity. As such they may be regarded as mild or at least weaker than the overall mean of the heterogenous stimulation previously experienced in life or experienced during the prolonged delay interval when the subject had left the laboratory. A central-tending effect or shift of sought value toward the mean of a series (5, pp. 56-57; 9, pp. 445-448) ought not to apply in the usual sense for these data, as each subject made but a single match with no opportunity to palp the range of a series. It might, however, be considered that, in the broader sense of a comparison of test stimuli with all sensory experience known to the subject, a central-tending effect could



come into play to enhance the value of the mild standard stimuli held through experimental delay. Neither the nature of the augmentation sometimes observed (e.g., of flash rate) nor the results of experiments exploring a systematic variation of standard strength (4) make this conclusion compelling, but it remains a possibility. There even may be a tending toward an intensification of the most identifiable stimulus characteristic in a *Gestalt*-like sense: e.g., that identified as *slow* made *slower*, *loud* made *louder*, and the like. As most of the questions relating to the possibly general tendency toward positive error in successive comparison over long delay are subject to experimental probing, they call for no extended speculation. Perhaps two general statements are justified on the basis of experimental results obtained thus far: (a) the evidence for the surprising accuracy with which delayed comparisons of the kind may be made points directly to the need for giving close attention to the details of the procedure used to demonstrate retaining and (b) the older naming of "time error," commonly associated with the method of successive-comparison, has never been clearly demonstrated to be a time error as such and the concept stands in need of a careful re-evaluation in terms of modern sensory psychophysics.

#### F. SUMMARY

The accuracy with which human subjects can reproduce sensory experiences of brightness, flash rate, loudness, pitch, and duration have been observed for delay intervals of 15, 30, and 60 seconds. Standard stimuli, presented under controlled conditions, were matched by the method of successive comparison by manual adjustment of variable stimuli following delay. The matches made to original (standard-stimulus) values were notably accurate and stable over all delay periods observed. When trends appeared in the constant errors obtained, they were consistent in both direction (sign) and degree with the errors observed over much longer delay intervals (one to 28 days) using identical test methods.

#### REFERENCES

1. KING, H. E. The retention of sensory experience: I. Intensity. *J. of Psychol.*, 1963, **56**, 283-290.
2. ———. The retention of sensory experience: II. Frequency. *J. of Psychol.*, 1963, **56**, 291-298.
3. ———. The retention of sensory experience: III. Duration. *J. of Psychol.*, 1963, **56**, 299-306.
4. ———. The retention of sensory experience: V. Variation of the standard stimuli (in preparation).
5. OSGOOD, C. Method and Theory in Experimental Psychology. New York: Oxford Univ. Press, 1953.

6. STARR, A., & LIVINGSTONE, R. B. Long-lasting nervous system responses to prolonged sound stimulation in waking cats. *J. Neurophysiol.*, 1963, **26**, 416-431.
7. STEVENS, S. S. Some similarities between hearing and seeing. *Laryngoscope*, 1958, **68**, 508-527.
8. ———. The surprising simplicity of sensory metrics. *Amer. Psychologist*, 1962, **17**, 29-39.
9. WOODWORTH, R. S. *Experimental Psychology*. New York: Holt, 1938.

*University of Pittsburgh*

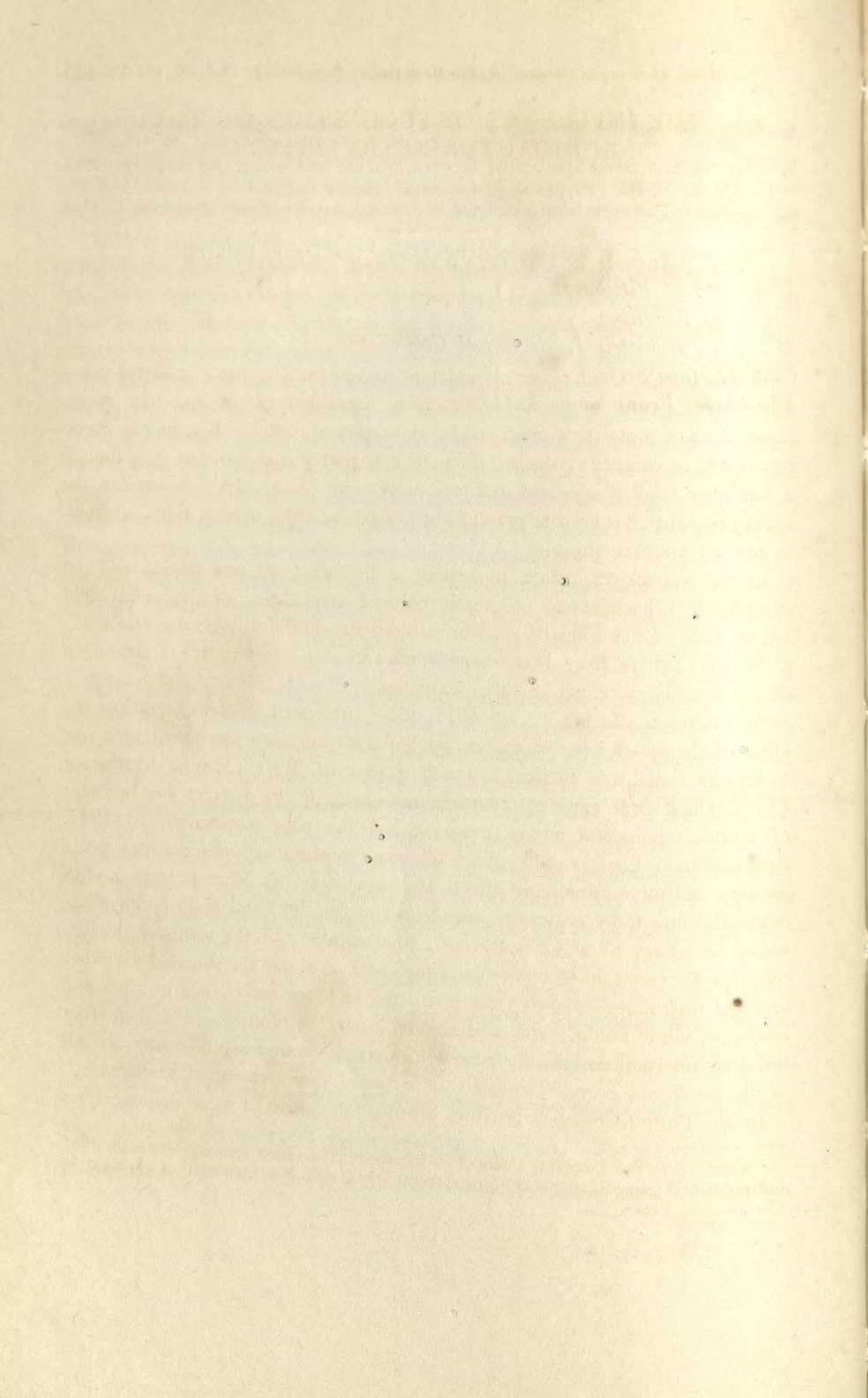
*School of Medicine*

*Western Psychiatric Institute and Clinic*

*3811 O'Hara Street*

*Pittsburgh, Pennsylvania 15213*





## THE INFLUENCE OF SUGGESTION ON THE AUDIOAUTOKINETIC EFFECT\*

*Psychology Department, Boston College*

---

JOSEPH R. CAUTELA AND DENNIS McLAUGHLIN

---

### A. INTRODUCTION

Wertheimer's classic paper on apparent movement (8) has stimulated many other investigators to perform laboratory investigations of apparent movement (i.e., of movement produced by stationary stimuli). Most of the investigators have concerned themselves with apparent visual movement, although a few have studied apparent auditory movement (3, 5). In these studies the usual procedure has been to present a stimulus and then, after a short interval, to present another stimulus.

In the autokinetic effect, movement is produced by one stationary-light stimulus in a background of greatly reduced articulation of the visual field (4, p. 156). Since Sherif's early investigations on the autokinetic effect (7, p. 95-105), there have been many studies concerned with the relationship of the autokinetic effect to other variables.

According to Boring (2, p. 601), those interested in demonstrating the unity of the senses have wanted to see apparent auditory movement turn out to be like visual and tactical apparent movement. Burt (3) and Mathiesen (5) obtained some apparent auditory movement by the discreet displacement of sounds, analogous to the procedure of the phi-phenomenon. In 1947, Bernadin and Gruber (1) presented some preliminary observations on a possible auditory autokinetic effect. In their study, Ss were presented with relatively pure tones (intensity not stated) of 56, 560, and 5600 cycles while seated in a dark room 12 feet from a loud speaker. In the series of instructions the Ss were told to report what they noticed about the sound. In another series of instructions, the Ss were instructed to listen for change in pitch and loudness; while nothing was said about changes in position. In still another series of instructions, the Ss were told that the sound would move. In all of the conditions every S reported some special displacement of the sound.

In our Boston College laboratory, during the course of some investigations

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 30, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



on sound localization, we observed as have many others [see (6)] that a sound source is difficult to locate when it is placed along the midline on the top of the head. It occurred to us that the foregoing phenomenon is analogous to the autokinetic effect in that adequate cues for localization are lacking.

In our preliminary investigations, we noticed that some *Ss* did not perceive auditory movement in all conditions of instructions. This fact was puzzling, in view of the results reported by Bernadin and Gruber (1); but our procedure involved two conditions not employed by Bernadin and Gruber. First, Bernadin and Gruber did not provide a head rest to fix the subject's head. We placed the *S's* head in a head rest, so that there would be little or no movement of the head. Second, we placed the sound source about *S's* head along the midline. Both of these conditions reduced localization cues. We felt that the conditions just specified made our audioautokinetic effect (AAE) more analogous than otherwise to the visual autokinetic effect (VAE).

Some of our preliminary studies indicated that *S* perceives the audioautokinetic effect to the extent that he is given suggestion on the likelihood of the movement of sound; so we decided to investigate systematically the effect of different instructions (degrees of suggestibility of movement) on the possible occurrence of AAE.

## B. METHOD

### 1. *Subjects*

The *Ss* were 40 male university students from 21 to 30 years of age.

### 2. *Apparatus*

The Auditory Autokinetic Apparatus [see Figure 1] consisted of an audio generator and a five-inch oxford speaker, Model No. 5C M-S, manufactured by the Oxford Company, Philadelphia, Pennsylvania. To the front of the speaker, the authors soldered a funnel, the openings of which were five inches at one end and one and three-quarters inch at the other. To the small openings, the authors soldered an 11-inch tube that served as a "gun barrel" with a bore of one-half inch. The entire apparatus was wrapped in rock wool, an insulator, gauze and covered with black tape. It was attached to a 10" × 9" rectangular box that served as a stand. The box was painted flat black.

### 3. *Procedure*

Each *S* was blindfolded outside the experimental room; then he was led into the experimental room and seated on a chair. His head was placed on a

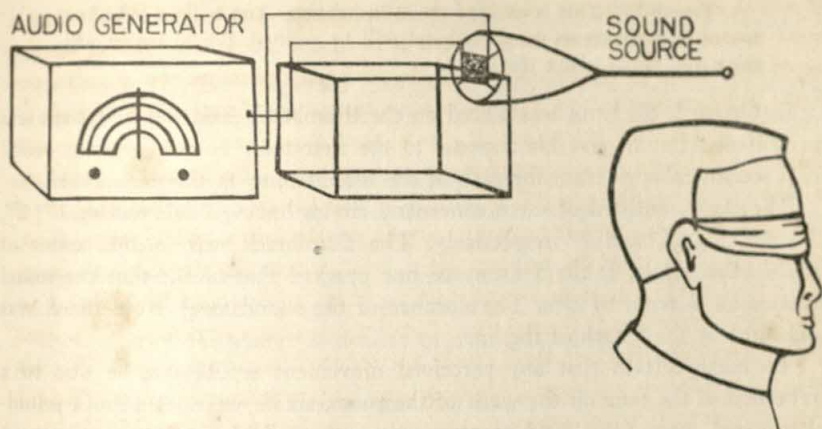


FIGURE 1  
THE APPARATUS USED TO PRODUCE THE AUDIOAUTOKINETIC EFFECT

head rest 12 1/2 inches above the edge of a table. The midline of the top of his head was determined by tape measure. After the midline was determined, the opening of the tube was placed three inches above S's head along the midline. Instructions were given, and a tone of 500 cycles and 35 db (sound-pressure level) was presented for two minutes. At the end of the two-minute period, the 550-cycle tone was presented again with an amplitude of 70 db. The instructions for the second tone were the same as for the first tone. S was asked to give his report at the end of each two-minute period.

Instructions for each of four experimental groups were as follows:

1. Group I. This is an experiment on auditory localization. You will be led blindfolded into a darkened room and be seated at a desk. Upon the desk will be a chart as seen below. At the center of the chart will be a thumbtack. This point will represent the center of the top of your head. A sound will be emitted at the top of your head first. You will be given a pencil and using the thumbtack as a starting point, your task will be to trace the movement of the sound.
2. Group II. This is an experiment in hearing. You will be led blindfolded into a darkened room and be seated at a desk. A sound will be emitted with different patterns of movement. I want you to describe the patterns of movement when I ask you to describe them.
3. Group III. This is an experiment in hearing. You will be led into a darkened room and be seated at a desk. A sound will be emitted. Please describe the characteristics of the sound when I ask you to describe them.



4. Group IV. This is an experiment in hearing. You will be led blindfolded in a darkened room. A sound will be emitted. Please report anything you notice about the sound.

In Group I, *S*'s hand was placed on the thumbtack, and the first tone was presented. After *S* gave his response to the first tone, he was given a pencil of a second color to trace the path of the second tone on the same chart.

The chart consisted of seven concentric circles having diameters of 1", 2", 3", 4", 5", 6", and 7" respectively. The thumbtack was in the center of the smallest circle. If the *S* drew the line upward that meant that the sound was going in front of him. The distance of the sound away from the *S* was indicated by the length of the line.

To make certain that any perceived movement would not be due to a reflection of the tone off the walls of the room, six *S*s were tested in a sound-attenuated room (this kind of room was not available at Boston College) under Group Instructions I, III, and IV.

After *S*s in Groups III and IV had given their responses to the two intensity levels, they were asked if they perceived any movement.

Every *S*, at the end of the experimental session (and while he was still blindfolded), was asked if he could locate the sound source in any part of the room.

### C. RESULTS AND DISCUSSION

Table 1 shows that *S*s reported movement when movement was suggested (Groups I and II). In Group I, each *S* started drawing the path of the

TABLE 1  
THE NUMBER OF *S*s REPORTING MOVEMENT

Group	Instruction	Intensity level			
		35 db		70 db	
		Yes	No	Yes	No
I	Trace	10	0	10	0
II	Describe	10	0	10	0
III	Characterize	1	9	5	5
IV	Notice	0	10	2	8

sound as soon as the sound source was turned on and continued to trace the path of the sound until the audiogenerator was turned off. Some *S*s stopped tracing for a second or two at various times during the experiment, but there was no difference in the amount or direction of movement according to intensity level. All the *S*s drew the path as far out as the third or fourth

circle. How far away each *S* perceived the sound source to be was limited, of course, by the size of the room ( $128'' \times 82''$ ).

In Group II, all *Ss* reported movement, with phrases such as "It started in the front and then went to the back," "The sound made a figure eight," "It went around my head a few times." Again there was no relation between the nature of the response and the intensity of the sound.

In Group III, five *Ss* reported movement as they described the characteristics of the sound. Four of the *Ss* reported movement with the 70-db intensity level only. One *S* reported movement with both intensities. The other characteristics reported usually were concerned with pitch and loudness or a description of the sound as similar to a familiar sound, such as a siren or the hum of a radio.

In Group IV, two *Ss* stated that the sound was moving; but reported movement only with the 70-db intensity level.

The foregoing results indicate that audioautokinetic movement is perceived when it is suggested. A chi-square analysis (Groups I and II combined *vs.* Group III and Group IV) yields an effect for instructions significant beyond the .01 level. The results also indicate that the more specific the instructions the greater the possibility of perceived movement. A similar trend was obtained with *Ss* in the sound-attenuated room. Two *Ss* traced the movement of the sound. Of the two *Ss* asked to report the characteristics of the sound, one perceived movement. Neither of the two *Ss* asked to report "what they noticed" perceived movement. Three *Ss* reported what Boring (2, p. 601) describes as the "phantom sound"; i.e., the sound moves around the head from one side to the other and either moves or jumps through the head to the other side and starts to move around again. These *Ss* also reported that they thought that the sound was coming from within their head.

After the *Ss* in Groups III and IV gave their responses to the two intensity levels, they were asked if they perceived any movement. In Group III, none of the *Ss* who had not reported movement previously answered "yes." All five *Ss* who previously reported that they perceived movement answered "yes." In Group IV, one *S* who previously did not say he perceived movement, reported movement. His response was "Now that you mention it, I think that I did." Two *Ss* who previously reported movement answered "yes"; the other *Ss* answered "no."

Of the 40 *Ss*, only one was sure that the sound source was overhead. Each of the other *Ss* expressed considerable amazement when the blindfold was taken off and the sound-source funnel was seen right over the top of his head.



Two significant conclusions are (a) that an audioautokinetic effect similar to the visual autokinetic effect is attainable and (b) the effect obtained is influenced by the degree of suggestion of the possibility of movement.

This characteristic of the audioautokinetic effect apparently is not applicable to the visual autokinetic effect. According to Sherif (7, 95-105), the visual autokinetic effect is easy to obtain and works without exception. Further, Sherif states that the effect takes place even when the *S* knows that the light does not move. A review of the literature on the visual autokinetic does not reveal any study similar to the present one investigating the extent to which he sees. Our own preliminary investigations reveal that perhaps the visual autokinetic effect is similar to the audioautokinetic effect in this respect; i.e., for most *Ss* it is necessary to suggest movement before they perceive the visual autokinetic effect.

Since there is an indication that the intensity level can influence the possibility of perceived movement, further experimentation is needed to determine the exact effects of systematic variation in intensity level.

#### REFERENCES

1. BERNADIN, A., & GRUBER, H. An auditory autokinetic effect. *Amer. J. Psychol.*, 1957, **70**, 133-134.
2. BORING, E. G. *Sensation and Perception*. New York: Appleton, 1942.
3. BURT, H. E. Auditory illusions of movement: A preliminary study. *J. Exper. Psychol.*, 1917, **2**, 63-75.
4. DEMBER, W. *Psychology of Perception*. New York: Holt, 1960.
5. MATHIESEN, A. Apparent movement in auditory perception. *Psychol. Monog.*, 1931, **41**, 94-131.
6. ROSENZWEIG, M. R. Development of research on the physiological mechanism of auditory localization. *Psychol. Bull.*, 1961, **58**, 376-389.
7. SHERIF, M. *The Psychology of Social Norms*. New York: Harper, 1936.
8. WERTHEIMER, M. Experimentelle studien über das Sehen von Bewegung. *Z. Psychol.*, 1912, **61**, 161-265.

*Psychology Department*

*Boston College*

*Chestnut Hill 67, Massachusetts*

## NEED FOR SOCIAL APPROVAL AND ATTITUDE CHANGE\*<sup>1</sup>

*Department of Psychology, Washington University*

ROBERT BUCKHOUT

### A. INTRODUCTION

In a previous experiment Buckhout and Rosenberg (2) found that verbal reinforcement of counterattitudinal statements was instrumental in producing varying amounts of attitude reorganization and change. These findings were predictable from propositions implicit in Rosenberg's structural theory of attitude dynamics [see Rosenberg (7) and Rosenberg *et al.* (8)].

The possibility that dramatic shifts in attitude might be due partially to situational compliance led the author to consider an experimental design that includes testing of a personality characteristic related to compliance behavior. Because need for social approval has been found to be related to the willingness of Ss to conform to perceived situational demands [see Marlowe and Crowne (5) and Crowne and Strickland (4)], the author decided to investigate the behavior of Ss who were high and low in need for social approval. He predicted that verbal reinforcement of the public verbalization of counterattitudinal statements would lead to attitude reorganization and change. He predicted further that Ss high in need for social approval are more likely to exhibit attitude change than are Ss low in need for social approval. A 30-day follow-up testing session was included to check on the relative "permanence" of the attitude changes produced.

### B. METHOD

#### 1. *Attitude Measures*

Three scales were employed to measure attitude toward television programming, an attitude object of relatively low salience.

*a. Long-term affect measure.* Each S rated his overall feeling toward television programming on a 16-point scale that ranged from -8 (extremely unfavorable) to +8 (extremely favorable).

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 30, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> Based on part of a dissertation submitted to the graduate school of The Ohio State University in partial fulfillment of the requirements for the Doctoral degree. The author acknowledges gratefully the guidance of Drs. Delos D. Wickens, Milton J. Rosenberg, Douglas P. Crowne, and Julian B. Rotter.



b. *Cognition score.* *E* presented to *Ss* a list of 12 value-like terms (i.e., relaxation, entertainment) representing goals to be gained from television programming and asked them to rate each term on a scale from -3 (highly undesirable) to +3 (highly desirable); also, on a scale from one to three, to rate the degree to which the goal is fostered by television programming. The two ratings were multiplied, and the algebraic sum of the product became the premanipulation cognition score.

c. *Intrainterview measures of conformity.* In the structured forced-choice interview described later, *Ss* (all of whom initially were *pro-television*) had the opportunity to verbalize consistent *pro-television* statements, despite the efforts of the interviewer, who verbally reinforced antitelevision or counterattitudinal statements. The frequency of counterattitudinal statements was used as an index of overall conformity. The difference between the frequency of counterattitudinal statements in the first half of the interview *vs.* that of those in the last half became an index of conformity shift.

## 2. Forced-Choice Interview

Under the pretext of gathering opinions on television programming, *E* gave *S* a deck of 44 cards containing two classes of paired statements about television programming. Twenty-eight of the cards contained a pair of antitelevision statements in a pair-comparison format. Because *S* was asked to choose and verbalize the statement closer to his own attitude, he was forced to verbalize a statement inconsistent with his initial (*pro-television*) attitude. *E* said "Good" regardless of *S*'s choice. Sixteen cards (test items) contained one *pro-television* statement and one antitelevision statement. *E* verbally reinforced *S* if *S* chose the antitelevision statement (i.e., conformed), but said nothing if *S* chose the *pro-television* statement.

## 3. Subjects

At the start of the semester, the Crowne-Marlowe Social Desirability Scale (3) was given to 300 female undergraduates, and initial affect-toward-television ratings were obtained. Fifty-four *pro-television* students were asked to serve as *Ss* and were promised credit points to fulfill a class-participation requirement.

## 4. Experimental Design

As *Ss* arrived, they were assigned to fill predetermined vacancies in a three-group design that included two groups of 18 *Ss* who were to be verbally reinforced while making choices during the forced-choice interview. One group was informed of the response-reinforcement contingency; the other

group was not informed. Each *S* who was informed was told that *E* would say "good" when he agreed with *S*'s choice and would say nothing when he did not agree. A third group of 18 *S*s acted as a control group, receiving no verbal reinforcement during the forced-choice interview. Half of the *S*s in each group were high and half were low in need for social approval, as measured by the Crowne-Marlowe Social Desirability Scale (mean of high *S*s = 21.2, mean of low *S*s = 12.4, mean of control *S*s = 16.7). *S*s in each cell were equated for initial affect toward television.

### 5. Procedure

*E* greeted *S*; gave him the list of value-like terms; collected ratings; computed the premanipulation cognition score; conducted the forced-choice interview; readministered the list of value-like terms; collected postmanipulation cognition scores; and gave *S* a questionnaire that called for a postmanipulation check on the long-term affect measure, *S*'s rating of *E*'s attitude toward television, and answers to questions on what *S* thought the purpose of the experiment to be. Approximately 30 days after the interview, *S*s were contacted and asked to complete a questionnaire that included the long-term affect scale, some buffer items, and a paper-and-pencil list of the forced-choice-interview items. *S*s were asked to choose the statement from each pair closer to their own attitude, but to complete the answer sheet in private.

### C. RESULTS

In general, both the *S*s who were informed and the *S*s who were not informed changed significantly in the expected direction on all measures ( $p < .05$ ), but no between-group differences were discovered; therefore in the subsequent analyses the experimental *S*s are combined under the need-for-social-approval categories: high and low. Table 1 shows that when affect

TABLE 1  
TRENDS IN LONG-TERM AFFECT AND COGNITION-CHANGE MEASURES

Mean	<i>S</i> s with a high need for social approval	<i>S</i> s with a low need for social approval	Control <i>S</i> s
Postinterview affect change	-1.61*	-0.89	-0.19*
30-day affect change	-1.00*	-0.34	0.06*
Cognition-change score	-3.81**	-0.50	3.12**

\* The starred means differ significantly from each other at the .025 level (Mann-Whitney one-tailed *U* test).

\*\* Starred means differ significantly from each other at the .001 level (Mann-Whitney one-tailed *U* test).



change is measured immediately after the interview, the high-need-for-social-approval Ss continue to show a significantly larger change from initial affect affect-change scores than do Ss in the control group (Mann-Whitney  $U$  test,  $p < .025$ ). The low-need-for-social-approval Ss do not differ significantly from control Ss. In the 30-day follow up, the high-need-for-social-approval Ss continue to show a significantly larger change from initial affect than do control Ss, despite the tendency of most Ss to show a slight turn toward initial affect levels.

Table 1 shows that high-need-for-social-approval Ss have a significantly larger cognition-score change in the predicted (i.e., reinforced) direction than do control Ss. Low-need-for-social-approval Ss show a slight (and nonsignificant) decline, while control Ss show an increase in cognition score.

The results for conformity shift are presented in Table 2 and show that high-need-for-social-approval Ss shift significantly more toward conformity than do control Ss. Low-need-for-social-approval Ss show a nonsignificant conformity shift.

TABLE 2  
TRENDS IN THE OVERALL-CONFORMITY SCORE AND THE CONFORMITY-SHIFT SCORE

Measure	Ss with a high need for social approval	Ss with a low need for social approval	Control Ss
Overall conformity during interview	12.1	9.2**	11.4
Overall conformity after 30 days	13.2*	8.4****	11.4
Mean conformity shift during interview	1.44***	0.67	0.06

\* At the .05 level (Mann-Whitney one-tailed test), the mean of 13.2 differs significantly from each other mean in the row.

\*\* At the .025 level, the mean of 9.2 differs significantly from the other means in the row.

\*\*\* At the .01 level, the starred mean differs significantly from the other means in the row.

\*\*\*\* At the .001 level, the mean of 8.4 differs significantly from the control-group mean of 11.4.

Table 2 also shows that on overall conformity the high-need-for-social-approval Ss do not differ from control Ss, but that the low-need-for-social-approval Ss choose significantly fewer counterattitudinal statements than do control Ss ( $p < .025$ ). Among experimental Ss, the number of counterattitudinal choices is positively correlated with the need for social approval ( $r = .38$ ,  $p < .05$ ). In the 30-day follow up, on a paper-and-pencil version of the interview scale, low-need-for-social-approval Ss have a significantly low

overall conformity score ( $p < .001$ ), while high-need-for-social-approval Ss have a significantly high overall conformity score. The changes over the 30 days, while not significant (according to Wilcoxon's test), are consistent in that control Ss show no change, high-need-for-social-approval Ss become *more* conforming, and low-need-for-social-approval Ss become *less* conforming. There are no significant differences between treatment groups in the ratings of *E*'s bias. *E* usually was felt to be slightly positive toward television. None of the Ss correctly reported the real purpose of the interview: namely, awareness. Most responses dealt with "television attitudes" and experimental details.

#### D. DISCUSSION

The procedures used were designed to produce attitude imbalance. The imbalance produced by verbally reinforced public verbalization of statements at variance with Ss' attitudes led to some dramatic attitude modifications. This effect, however, is not directly analogous to that in the typical verbal-conditioning study involving the "operant learning" of a reinforced-response class because the forced-choice scale was constructed so that Ss could not help but give a response from the reinforced-response class. All experimental Ss made at least 28 verbally reinforced "conforming" choices. On the 16 test items experimental Ss started to make consistent choices, but quickly showed a significant shift toward the reinforced counterattitudinal choice. The cognition scores just before and just after the interview showed a similar trend.

Despite negative findings by Spielberger *et al.* (9), the present experiment confirms the idea that Ss high in need for social approval tend to modify their behavior to obtain the social reinforcement provided in the experimental interview situation. On all measures, the high-need-for-social-approval Ss changed their attitudes more than did the low-need-for-social-approval Ss. Attitude change generally was in the expected negative direction under verbal reinforcement and tended to be in the positive direction in the control condition.

The several attitude-change measures can be evaluated in terms of their proximity to the attempt at persuasion (reinforcement). Those measures (cognition change and conformity scores) that were strictly part of the experimental session showed relatively more change in the reinforced direction than did the long-term affect-change measure. This result indicates that a person who is high in need for social approval conforms to the immediate sources of social influence. The question remaining is whether or not the attitude changes observed in most communication and persuasion studies repre-



sent basic changes in attitude or mere conformity to situational demands. The follow-up testing in the present study gives some evidence that the effects of the experimental manipulation lasted at least 30 days, especially for the high-need-for-social-approval Ss; therefore people who are high in need for social approval are more likely to conform and show attitude change (immediate and long term in nature).

#### E. SUMMARY

Ss were given a forced-choice scale during an interview that forced them to verbalize publicly inconsistent attitudinal statements in the context of positive verbal reinforcement. The procedures led to significant attitude change and conformity behavior among verbally reinforced Ss who were high in need for social approval. Some attitude-change effect remained after 30 days in high-need-for-social-approval Ss. It is concluded that need for social approval is an important source of motivation in the prediction of attitude change.

#### REFERENCES

1. BUCKHOUT, R. Verbal reinforcement and verbal behavior. An unpublished Doctoral dissertation. The Ohio State University, Columbus, Ohio, 1962.
2. BUCKHOUT, R., & ROSENBERG, M. J. Verbal reinforcement and attitude change. 1964 (In preparation).
3. CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *J. Consult. Psychol.*, 1960, **24**, 349-354.
4. CROWNE, D. P., & STRICKLAND, B. R. The conditioning of verbal behavior as a function of the need for social approval. *J. Abn. & Soc. Psychol.*, 1961, **63**, 395-401.
5. MARLOWE, D., & CROWNE, D. P. Social desirability and response to perceived situational demands. *J. Consult. Psychol.*, 1961, **25**, 109-115.
6. MOOS, R. H. The retention and generalization of operant conditioning effects in an interview situation. *J. Abn. & Soc. Psychol.*, 1963, **66**, 52-58.
7. ROSENBERG, M. J. A structural theory of attitude dynamics. *Pub. Opin. Quart.*, 1960, **24**, 319-340.
8. ROSENBERG, M. J., et al. Attitude organization and change. New Haven: Yale University Press, 1960.
9. SPIELBERGER, C. D., BERGER, A., & HOWARD, K. Conditioning of verbal behavior as a function of awareness, need for social approval and motivation to receive reinforcement. *J. Abn. & Soc. Psychol.*, 1963, **67**, 241-246.

Department of Psychology  
Washington University  
St. Louis, Missouri 63130

## PERSONALITY CHARACTERISTICS, CIGARETTE SMOKING, AND OBSTETRIC COMPLICATIONS\*

*Department of Psychiatry, Emory University School of Medicine*

---

ROBERT L. McDONALD

---

### A. INTRODUCTION

Observations that smokers differ from nonsmokers in various aspects of personality have stimulated considerable research focusing on the relationship between personality factors and smoking habits. Smokers have been found to be more impulsive (13), more "chance" oriented (16), less suggestible (17), more anxious and have a greater number of psychosomatic symptoms (8), more neurotic (6), and to have higher mean psychopathic-deviate and hypomania but lower social-introversion MMPI scores than nonsmoking counterparts (12). In addition, Eysenck (2) has reported linear relationships between the amount of smoking and extraversion and rigidity.

The effects of smoking on parturient females have also been demonstrated; that is, the mean birth weight of smokers' offspring is lower than that of the offspring of nonsmoking mothers (3, 7, 15). In terms of manifest anxiety, neuroticism, and the MMPI neurotic and psychotic scales (10, 11, 14), further studies have differentiated patients with prematurity, as well as with other obstetric complications, from patients with uncomplicated pregnancies and deliveries. These results, collectively, imply the possibility that common personality factors (e.g., high manifest anxiety) underlie smoking and obstetric complications, but the literature reveals no systematic attempt to study the foregoing variables as conjoined functions of personality. Therefore, in the present study, the author tests the hypothesis that smokers with obstetric complications have personality characteristics significantly different from those of nonsmokers with uncomplicated pregnancies.

### B. METHOD

One hundred twenty-nine white unmarried primigravidas (first-pregnancy patients) attended at the Medical College of South Carolina were first cate-

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on May 3, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



gorized (following delivery) as uncomplicated or as complicated, depending on the presence or absence of a variety of common obstetric complications observed during pregnancy or delivery. [See McDonald and Parham (11) for details.] Next, subjects were trichotomized as nonsmokers, light smokers (up to 10 cigarettes—i.e., one-half package daily), and heavy smokers (more than 10 cigarettes daily). Ages ranged from 13 to 31 ( $M = 18.6$ ) and showed no relation to smoking habits or obstetric complications. The majority of the subjects were of lower middle-class socioeconomic status.

The Minnesota Multiphasic Personality Inventory was administered in group form during the third trimester (i.e., at the beginning of the 8th month of gestation). At time of testing, each subject completed a personal-data sheet that asked for data on smoking habits. Any smoker who changed her smoking habits during pregnancy was dropped from the study. There were five such cases.

### C. RESULTS

The major findings of this study are presented in Tables 1 and 2. Table 1 shows significant  $F$  values for smoking habits in relation to the depression, psychopathic-deviate, paranoia, psychasthenia, schizophrenia, and hypomania scales; but no significant values for obstetric complications or interaction effects.<sup>1</sup>

Table 2 shows that with two exceptions all mean scores are within normal limits, and most groups peak on the psychopathic-deviate scale. Heavy smokers were more depressed than light smokers ( $t = 2.40$ ,  $p < .02$ ) and tended to be more depressed than nonsmokers ( $t = 1.98$ ,  $p > .05$ ). On the psychopathic-deviate scale heavy smokers had higher scores than nonsmokers ( $t = 2.48$ ,  $p < .02$ ). On the paranoia scale, nonsmokers obtained higher mean scores than light smokers ( $t = 2.68$ ,  $p < .01$ ). Light smokers obtained lower psychasthenia scores than either nonsmokers or heavy smokers ( $t = 2.57$ ,  $p < .02$ ;  $t = 2.20$ ,  $p < .05$ ). On the schizophrenia scale both nonsmokers and heavy smokers obtained higher scores than light smokers ( $t = 2.60$ ,  $p < .02$ ;  $t = 2.17$ ,  $p < .05$ ). On the hypomania scale nonsmokers and light smokers obtained higher scores than heavy smokers ( $t = 2.04$ ,  $p < .05$ ;  $t = 2.85$ ,  $p < .01$ ).

<sup>1</sup> Analysis of variance tables have been deposited with the American Documentation Institute, Order Document No. 8401 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D.C. Remit in advance \$1.75 for microfilm or \$2.50 for photocopies, and make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1  
F VALUES FROM ANALYSES OF VARIANCE OF MMPI SCALES BY SMOKING AND OBSTETRIC COMPLICATIONS

F VALUES FROM ANALYSES OF VARIANCE OF MINIM SCALE 24														
Source	L	F	K	Hs	D	Scale*		Pd	Mf	Pa	Pt	Sc	Ma	Si
Smoking (A)	.09	1.28	.26	2.13	3.66**	2.04	3.14**	.87	3.70**	4.08**	3.52**	3.88**	2.85	
Complications (B)	.05	.04	.06	.61	.01	.02	.25	1.40	.29	.00	.15	.78	.03	
A × B	1.68	.06	.41	.99	.67	1.67	.47	1.04	.79	.69	.07	1.30	1.80	

\* See Table 2 for titles of scales.

\*\*  $p < .05$ .



TABLE 2  
MMPI SCALE MEANS AND STANDARD DEVIATIONS BY SMOKING  
AND OBSTETRIC COMPLICATIONS

Scale	Non-smoker (N = 23)	Normal Light Smoker (N = 14)	Heavy Smoker (N = 11)	Non-smoker (N = 41)	Complicated Light Smoker (N = 16)	Heavy Smoker (N = 24)
Lie (L)	49.26 7.72	50.00 8.39	47.55 5.22	49.17 8.34	47.50 6.14	51.45 7.01
Validity (F)	58.17 10.41	55.14 9.66	58.82 9.35	59.54 12.44	55.44 8.97	58.42 8.83
Test-taking attitude (K)	51.39 10.61	54.79 8.76	51.73 10.92	52.64 11.35	51.94 10.01	51.88 6.49
Hypochondriasis	57.09 13.59	52.50 10.68	52.09 7.47	57.27 9.58	51.50 10.23	57.71 10.43
Depression	61.57 11.65	53.71 12.19	60.09 13.42	57.71 10.36	54.50 12.41	62.33 10.43
Hysteria	59.61 12.35	55.64 8.27	56.00 9.09	57.54 8.53	53.38 10.07	61.21 10.19
Psychopathic- Deviate	69.39 13.88	66.64 9.64	73.09* 12.11	64.44 10.07	67.56 10.41	72.88* 10.84
Masculinity- femininity	50.26 7.12	48.93 7.58	53.64 7.16	51.51 9.76	53.94 6.81	53.08 8.51
Paranoia	64.96 15.20	53.50 11.27	61.18 11.82	60.98 10.01	56.44 12.86	58.33 13.94
Psychasthenia	63.65 13.38	54.93 9.78	64.36 14.78	61.98 8.85	58.50 10.27	62.54 10.91
Schizophrenia	66.57 16.45	59.93 10.01	66.91 13.08	67.00 13.98	58.00 10.97	65.29 14.54
Hypomania	61.70 11.93	62.29 10.58	59.09 11.08	63.10 11.88	68.93 12.96	56.96 10.23
Social Introversion	60.26 8.97	51.86 9.12	57.18 13.33	56.66 10.51	53.19 8.58	58.42 8.56

\* Except for the means marked with an asterisk, all means are within normal limits.

#### D. DISCUSSION

The hypothesis that heavy smokers with obstetric complications differ in personality characteristics from nonsmokers without complications is not supported; moreover, no significant differences were found on the obstetric-complications variable.

A possible explanation for the lack of positive results may be the particular point in time the test data were obtained. In previous studies (9, 11), the author has found reliable MMPI scale differences at the seventh month of pregnancy between subjects with obstetric complications and those with normal pregnancies. Further, McDonald and Parham (11) have reported the virtual disappearance of group-personality differences postpartum. One won-

ders, therefore, if there is a denotable time after which personality differences between groups of patients with obstetric complications and without begin to disappear.

The data presented do not establish a consistent pattern, but they imply a curvilinear relationship between smoking habits and personality characteristics. On the depression, paranoia, psychasthenia, and schizophrenia scales, mean scores for light smokers are lower than those of either the nonsmokers or heavy smokers. A U-shaped relation seems inconsistent with Eysenck's findings of linear relationships between smoking and rigidity and extraversion (2) despite the fact that rigidity and extraversion are dissimilar to the symptom-oriented MMPI scales. The present author's results suggest less maladjustment than otherwise for light smokers, a suggestion tested by L'Abate's scatter measure (4). An analysis of variance, using L'Abate's measure (a global index of maladjustment derived by summing the absolute deviations from the standard means for each of the nine clinical scales), supports the assertion that light smokers show fewer symptoms than do the nonsmokers or heavy smokers.

Several hypotheses relating smoking habits to MMPI scores have been previously formulated. For example, Schubert (12) hypothesized and demonstrated that depression varies directly with smoking. Despite the foregoing, the present author found that heavy smokers obtained reliably lower hypomania and higher depression scores than either light smokers or nonsmokers. The interpersonal interpretation of these results [see (5)] is that heavy smokers emphasize docile-dependent behavior and understress competitive-narcissistic behavior. When compared with that of nonsmokers, the psychopathic-deviate scores of smokers generally has been found to be high. Our results corroborate this finding, but again do not support a linear relationship between smoking and this personality variable.

Several theories, ranging from the common-sense hypotheses of anxiety and oral indulgence to arousal seeking (13) and Bales' dual needs and social-acceptance theory (1), have been put forward to explain smoking. Our results are inconsistent with the oral-indulgence theory, and analysis of variance using available manifest anxiety scores fails to support the anxiety theory.

#### E. SUMMARY

In this study the author studied the relation between smoking habits and obstetric complications on MMPI scores obtained at the 8th month of pregnancy from unmarried white primigravidas. Significant relations between



smoking habits and scores on the depression, psychopathic-deviate, paranoia, psychasthenia, schizophrenia, and hypomania scales were found. These data were discussed in the light of previous findings and current theories.

## REFERENCES

1. BALES, R. F. Social therapy for a social disorder—compulsive drinking. *J. Soc. Issues*, 1945, **23**, 1-9.
2. EYSENCK, H. J., TARRANT, M., WOOLF, M., & ENGLAND, L. Smoking and personality. *Brit. Med. J.*, 1960, **7**, 1456-1460.
3. FRAZIER, T. M., DAVIDS, G. H., GOLDSTEIN, H., & GOLDBERG, I. D. Cigarette smoking and prematurity; a prospective study. *Amer. J. Obst. & Gynecol.*, 1961, **81**, 988-996.
4. L'ABATE, L. MMPI scatter as a single index of maladjustment. *J. Clin. Psychol.*, 1962, **18**, 142-143.
5. LEARY, T. F. Multilevel Measurement of Interpersonal Behavior. Berkeley: Psychological Consultation Service, 1956.
6. LILIENTHAL, A. M. Emotional and other selected characteristics of cigarette smokers and nonsmokers as related to epidemiological studies of lung cancer and other diseases. *J. Nat. Cancer Inst.*, 1959, **22**, 259-282.
7. LOWE, C. R. Effect of mothers' smoking habits on birth weight of their children. *Brit. Med. J.*, 1959, **2**, 673-676.
8. MATARAZZO, J. D., GUSE, S., & MATARAZZO, R. G. An approach to the validity of the Taylor anxiety scale; Scores of medical and psychiatric patients. *J. Abn. & Soc. Psychol.*, 1955, **51**, 276-280.
9. McDONALD, R. L., & CHRISTAKOS, A. C. Relationship of emotional adjustment during pregnancy to obstetric complications. *Amer. J. Obst. & Gynecol.*, 1963, **86**, 341-348.
10. McDONALD, R. L., GYNTHY, M. D., & CHRISTAKOS, A. C. Relations between maternal anxiety and obstetric complications. *Psychosom. Med.*, 1963, **25**, 357-363.
11. McDONALD, R. L., & PARHAM, K. J. Relation of emotional changes during pregnancy to obstetric complications in unmarried primigravidae. *Amer. J. Obst. & Gynecol.*, 1964, **90**, 195-201.
12. SCHUBERT, D. S. P. Personality implications of cigarette smoking among college students. *J. Consult. Psychol.*, 1959, **23**, 376.
13. ———. Arousal seeking as a central factor in tobacco smoking among college students. *Int. J. Soc. Psychiat.*, (in press).
14. SCOTT, E. M., & THOMSON, A. M. A psychological investigation of primigravidae. IV. Psychological factors and the clinical phenomena of labor. *J. Obst. & Gynecol. Brit. Emp.*, 1956, **63**, 502-508.
15. SIMPSON, W. J. A preliminary report on cigarette smoking and the incidence of prematurity. *Amer. J. Obst. & Gynecol.*, 1957, **73**, 808-815.
16. STRAITS, B. C., & SECHREST, L. Further support of some findings about the characteristics of smokers and nonsmokers. *J. Consult. Psychol.*, 1963, **27**, 282.
17. VALLANCE, T. R. Suggestibility of smokers and nonsmokers. *Psychol. Rev.*, 1940, **4**, 138-144.

Department of Psychiatry  
 School of Medicine  
 Emory University  
 Atlanta, Georgia 30322

## THE PROBLEM OF BACKWARD CONDITIONING\*

*Department of Psychology, Boston College*

---

JOSEPH R. CAUTELA

---

### A. INTRODUCTION

Even though the process of backward conditioning is a crucial problem for some learning theorists (21, 31) and an important issue in its own right as part of the integral empirical body of knowledge concerning conditioning, the validity of backward conditioning as a genuine conditioning phenomenon remains in dispute.

Pavlov (25, pp. 27, 393) was the first to deny the possibility of backward conditioning, and he did this on the basis of Krestovnikov's experiments (20). Pavlov later retracted this position (26, p. 381) and said that even though a conditioned reflex may develop if the conditioned stimulus is introduced after the unconditioned stimulus his experimentation indicated the conditioned reflex that would develop is insignificant and evanescent, and on continuing the trials the conditioned stimulus becomes inhibitory.

In 1956 Razran, in an article reviewing the nature of backward conditioning with special emphasis on Russian studies (31), concluded that backward conditioning is a genuine conditioning phenomenon. In spite of Razran's analysis and conclusions, some current commentators on conditioning continue to assert that the existence of backward conditioning as a genuine conditioning phenomenon is doubtful. Brogden (6, pp. 579-580) and Kimble (18, pp. 158-159) hold that any backward conditioning obtained is probably due to pseudo-conditioning. Osgood (24, pp. 313-314), Woodworth and Shlosberg (39, pp. 567-568), and Deese (10, pp. 43-44) argue that reported cases of backward conditioning are the result of sensitization.

### B. PURPOSE

Why this disagreement between Razran and other investigators? Usually a disagreement of this kind is due to a lack of agreed-upon criteria for the phenomenon under discussion or a difference in interpretation of results. This

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on May 4, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



article is an attempt to analyze the evidence and criteria of backward conditioning and to clarify the problem.

One of the factors that has led to confusion concerning backward-conditioning phenomena is the definition of backward conditioning employed by various investigators. Of the many attempts to define conditioning, two definitions have become prominent. One definition can be called the substitute-stimulus definition. The substitution definition is essentially that of Hilgard (17, p. 50). A neutral stimulus is paired with a stimulus that always elicits Response A. If after a number of pairings the heretofore neutral stimulus elicits Response A, conditioning is said to have taken place. The modifying-stimulus definition is presented by Barlow (1, p. 406) as the process of modifying the effects of a stimulus on the behavior of organisms by associating it with a second stimulus. It will be seen later that operant-conditioning studies on backward conditioning use this definition.

The substitute-stimulus definition implies that the response to the conditioned stimulus must be similar to the response to the unconditioned stimulus. In the modifying-stimulus definition, the conditioned stimulus need not elicit a conditioned response similar to the one elicited by the unconditioned stimulus. Also, according to the modifying-stimulus definition, the stimulus can inhibit as well as facilitate response probability.

Most investigators who hold the substitute-stimulus definition of conditioning also hold that the conditioned-response probability must increase with repeated reinforcements and remain relatively stable when some maximum strength is reached. According to this viewpoint, unstable or temporary conditioned responses should not be classified as true conditioning. From this viewpoint, the unstable or temporary conditioned responses would be attributed to pseudoconditioning or sensitization. Pseudoconditioning is defined as "The strengthening of a response to a previously neutral stimulus through repeated elicitation of the response by another stimulus without paired presentation of the two stimuli" (18, p. 482). Sensitization is defined as "the increase of a reflex originally evoked by a conditioned stimulus through its conjunction with an unconditioned stimulus and response" (8, p. 483).

With the foregoing definitions of conditioning and their implications in mind, a review and categorization of the backward-conditioning studies is in order.

### C. RUSSIAN STUDIES

In his review of the Russian studies concerning backward conditioning, Razran states "On the whole, the analyzed evidence is unmistakable in demonstrating that B.C. is not a case of pseudoconditioning, but is a genuine CR

—associative manifestation, and the stable backward CR's can be obtained and maintained under favorable experimental conditions" (31, p. 67).

A close look at the Russian studies presented does not give the present author unmistakable evidence that according to the substitute-stimulus definition (accepted by Razran as the only appropriate definition) stable and well developed backward-conditioning responses have taken place.

Razran discusses the Russian studies in chronological order. He states that Krestovnikov (20) really did show backward conditioning in two dogs similar to the conditioned responses obtained by Wolffe (37, 38). The conditioned responses were unstable and tended to diminish with more reinforcement.

In the experiments reported in 1933, Kreps (19) found unstable conditioned responses. Pavlova (27) had found some backward-conditioned responses but not as effective as forward-conditioned responses. Vinogradov (36) found backward conditioning, but the conditioned responses that developed tended to diminish with repeated reinforcements. Petrova (28) found small conditioned responses that tended to disappear in the course of training. The 1933 studies show that backward-conditioned responses can be obtained, but they are unstable and tend to diminish with training.

In experiments reported in 1940, Razran cites the only case (Nezhdanova) in all the Russian and the American literature, so far as the author can determine, in which true conditioning (exhibiting the properties of a forward-conditioned response) was obtained in a classical-conditioning paradigm. Nezhdanova (23) used a 0.3 per cent solution of HCl as the unconditioned stimulus and bubbling water for a conditioned stimulus. Backward-time intervals between three and five seconds were used. The conditioned response to the backward-conditioned stimuli appeared after the 12th reinforcement and continued to increase in magnitude and showed no decrease after 445 reinforcements. In all respects, the conditioned responses behaved like the well-established forward-conditioned responses. In other 1940 studies (22, 34), little or no backward conditioning was established. In other miscellaneous studies presented by Razran, Beritov (3, 4) and Pressman (30) did not obtain stable backward conditioning. An overall examination of the studies presented by Razran reveals that in all but one of the studies of backward conditioning, the conditioned responses that did occur were of the unstable variety or diminished with training. The evidence is not as unmistakable as Razran would have one believe.

#### D. AMERICAN STUDIES

American studies on backward conditioning have been of the classical-conditioning and operant-conditioning variety. In the operant-conditioning



type of study, the modifying-stimulus definition of conditioning has been used. Razran (31, p. 56) dismisses the operant-conditioning type of study as a valid test of backward conditioning: "Since in operant conditioning the conditioned reaction by definition produces the unconditioned one and obviously precedes it" (31, p. 56). But writers investigating backward conditioning with operant-conditioning procedures claim that they have fulfilled the definition of backward conditioning by presenting a conditioning stimulus after the reinforcement. They test the conditioned stimulus to see if it has picked up reinforcing properties. In other words, they apply the modifying-stimulus definition of conditioning to backward conditioning.

### 1. *Classical-Conditioning Studies*

In American studies employing classical-conditioning procedures, the first significant classical-conditioning studies on backward conditioning are the well-known Wolffe experiments (37, 38). She used sound as a conditioned stimulus and shock as an unconditioned stimulus. She used backward intervals of .25 second and .50 second. Her results revealed 13 per cent and 10 per cent conditioning respectively. In the second experiment (38) she used backward intervals of .2, 1.6, 1.0 and 2.0 seconds. The mean percentage conditioning for four intervals were seven, 10, 11, and six per cent respectively. She also found that one-second-backward conditioning was superior to one-second-forward conditioning in per cent conditioning. Spooner and Kellogg (33) were to comment later that pseudoconditioning was probably involved in the Wolffe experiments.

About the same time, Switzer (35) employed backward-conditioning intervals from 0.5 to one second. He used mild shock as an unconditioned stimulus and a "click" as the conditioned stimulus. He reported that 18 out of 20 Ss (90 per cent) showed backward conditioning; but the stability varied markedly among individuals, and the amplitude of the conditioned response decreased as training progressed.

In 1934, Bernstein (5) used backward intervals of 0.5 and .90 second. An eyelid reaction to shock was paired with a "click." Only seven out of 20 subjects showed backward conditioning, and the greatest amount of conditioning in a subject was eight per cent. He attributed the slight amount of conditioning he and Switzer (35) obtained as due to pseudoconditioning.

Cason (8), using natural eyelid reaction as the unconditioned stimulus and a "click" as the conditioned stimulus found no backward conditioning in eight

human Ss when the conditioned stimulus was presented after the eyelid reaction was four-fifths complete.

In a study by Grether (15) using monkeys as Ss, no difference was found in the acquisition of a conditioned response produced by backward and pseudoconditioning procedures.

Porter (29) reported no backward conditioning of the eyelid response in human Ss. The unconditioned stimulus was a puff of air, the conditioned stimulus was a flash of light, and intervals of 470 and 980 milliseconds were used.

Harris's study (16), using shock to the finger as an unconditioned stimulus followed immediately by a tone, obtained some backward conditioning; but he attributed it to pseudoconditioning factors rather than to temporal factors.

In Spooner and Kellogg's study (33), finger withdrawal to shock was the unconditioned response; and finger withdrawal to tone was the conditioned response. Backward intervals of 0.05 second and 0.25 second were used. There was also a simultaneous group and a forward group of 0.50 second, 1.00 second, and 1.50 seconds. Spooner and Kellogg found that the conditioned responses in the forward-conditioning group showed a typical sigmoid conditioning curve. The curve for simultaneous and backward conditioning showed a process of decrement or extinction, which is apparently the reverse of simultaneous and forward conditioning. In other words, as the number of conditioning trials increases, the conditioned responses for the backward-conditioning group become less frequent. There was also an important difference in the latency scores of the forward-conditioning groups as compared with those for the simultaneous-conditioning and backward-conditioning groups. The latency of the forward-conditioned responses tended to increase with practice, while this progression was not observed with the backward-conditioning groups. The intensity of the reinforcing shock needed to keep the unconditioned response approximately constant in magnitude gradually tended to increase for backward-conditioning groups, but remained the same for the forward-conditioning groups. Spooner and Kellogg conclude (33, p. 344) that the differences in frequency and latency of the conditioned responses and differences in intensity of the unconditioned stimulus point to the fact that forward conditioning and backward conditioning are fundamentally distinct processes, and the data of the two processes should not be treated together in the same category. They conclude that backward-conditioned responses probably are special cases of pseudoconditioning.

In a more recent study Fitzwater and Reisman (11), using tone as the



conditioned stimulus and finger withdrawal to shock as the unconditioned response, found that a forward interval of 0.5 second led to an unstable formation of the conditioned response. The frequency of conditioned responses during training showed an increase in early training, then declined with further training.

In all of the preceding studies, backward conditioning was not obtained or was attributed to pseudoconditioning (by the investigators or evaluated as such by others).

## 2. *Operant-Conditioning Studies*

In studies using operant-conditioning procedures as methods of investigating backward conditioning, the modifying-stimulus definition has been used.

Carr and Freeman (7) found that nine white rats failed to learn in 1,500 trials to turn around and retrace a path at the sound of a buzzer when it was sounded one second after the door of the path that led to food was closed.

Yarborough (40) gave rats a shock immediately followed by a buzzer when an animal had to turn and retrace a pathway in a maze. Progress was measured by the increasing percentage of correct responses in test trials. The backward-conditioning groups showed learning like the forward-conditioning group with a two-second interval.

More recently, Coppoch (9) reported that a light blink picked up reinforcing properties if it was presented two seconds after the termination of shock.

Barlow (1, 2) presented a light stimulus immediately after a 10-second shock. His results showed a greater bar-pressing response acquired by this group than another with shock alone and no-shock groups when the light stimulus was presented.

Smith and Buchanan (32) found that animals trained with a black alley following shock made fewer errors and took fewer trials to reach a criterion (black positive) on a black-white choice situation than rats trained in a white alley not following shock.

Finally, Goodson and Brownstein (14) observed that rats allowed to escape shock from a black box by going into a white box made significantly more choices to a box similar to the escape chamber than they did to a neutral box.

## E. CONCLUSIONS

The following conclusions can be derived.

1. Only one study seems unquestionably to have fulfilled the criteria for

the substitute-stimulus definition of conditioning. In all other studies, the conditioned responses were unstable, did not become stronger with practice, or were called pseudoconditioning by some investigator.

2. Some studies (1, 9, 14, 32, 40) seemed to have fulfilled the criteria for the modifying-stimulus definition of conditioning.

3. In every study fulfilling the criteria of conditioning, a noxious stimulus was used as the unconditioned or reinforcing stimulus. In the only classical-conditioning study in which backward conditioning seemed a genuine phenomenon, Nezhdanova used HCl as the unconditioned stimulus; in the operant-conditioning studies, the unconditioned stimulus was a shock. It seems that Switzer (35, p. 89) was right when he said that backward conditioning occurs most readily when defense reactions are involved. The apparent necessary involvement of a noxious stimulus in backward conditioning supports the sensitization explanation. Representing this point of view, Deese (10, p. 43-44) is of the opinion that in backward-conditioning studies in which electric shock was used the *S* is so sensitized that he will respond to almost any stimulus.

#### F. IS BACKWARD CONDITIONING A GENUINE PHENOMENON?

If the modifying-stimulus definition of conditioning is used, a number of studies fulfill the definition. If the substitute definition is used, only one study seems to fulfill all the criteria. All the studies in which backward conditioning has occurred employed a noxious stimulus. This still leaves the possibility that the backward-conditioning phenomenon is due to sensitization. The pseudoconditioning explanation is not adequate to explain all reported cases of backward conditioning. Razran (31, p. 59) stated that pseudoconditioning could not be an explanation of the backward conditioning obtained in Russian studies since in no case were the conditioned responses elicited by stimuli that had not been paired with the conditioned stimuli. In other words, pseudoconditioning procedure was not used.

#### G. AN ALTERNATE HYPOTHESIS

An alternate hypothesis to the sensitization explanation of why backward conditioning has been obtained only with a noxious stimulus concerns the duration of perceived pain. According to this hypothesis, cases of reported backward conditioning with the use of a noxious stimulus are really cases of the simultaneous conditioning of the conditioned stimulus and pain perception.



One of the noxious-stimuli characteristics that results in pain perception (as determined by voluntary withdrawal of the affected part) is the perception of pain for some time after the noxious stimulus is removed. A pinch on the hand or a blow on the arm usually will result in the perception of pain (if the stimulus is intense enough) after the noxious stimulus is removed. All other things being equal, decrease in the intensity of the pain perceived is directly related to the amount of time since the application of the noxious stimulus. The two kinds of noxious stimuli used in successful backward-conditioning studies have been a 0.3 per cent solution of HCl (23) and electric shock. In humans, the shock usually is applied to the hand or finger; in animals, to the paws. The studies reporting backward conditioning indicate the closer the conditioned stimulus is to the unconditioned stimulus the more effective the conditioning. Most successful backward conditioning occurs in a matter of seconds. There is some evidence from the laboratory at Boston College that pain perception, at the threshold of withdrawal, can last as long as one minute after shock is delivered to the hand. The threshold levels for hand withdrawal were determined for 20 male undergraduates. The shock was delivered to the right hand, and the Ss were instructed to press a buzzer with their left hand when they no longer felt pain in the right hand. The response times varied from five seconds to one minute. These findings, of course, are complicated by the individual's own subjective definition of pain.

There is further evidence to support the likelihood that pain perception to shock can last for more than a minute after the shock stimulus is removed. Fulton (12, p. 344) and Gardner (13, p. 163) report that A fibers (i.e., larger myelinated fibers) conduct rapidly the intense sharp pain followed a second or two later by a delayed pain that tends to radiate and is extremely unpleasant. The delayed pain is conducted by the very small, unmyelinated C fibers. The delayed pain can last a matter of minutes after the noxious stimulus is removed even if there is no tissue damage. Of course, anyone who has participated in an experiment in which finger or hand withdrawal to shock was involved will report the pain is sharp and intense (and does not cause tissue damage).

Though at this stage of development the evidence supporting the duration-of-pain hypothesis is admittedly meager, it is preferable to the sensitization hypothesis because it is more specific. The sensitization hypothesis merely asserts that the noxious stimulus somehow heightens the probability that the organism will respond to any well-defined stimulus in a manner similar to the response to the noxious stimulus. The organism is somehow alerted, triggered, or primed for any stimulus that occurs temporally near the noxious

stimulus. Actually the sensitization explanation is a description of events. It does not attempt to explain why the triggering or priming occurs. The duration-of-pain hypothesis asserts that the conditioned responses occur in a situation in which the external unconditioned stimulus is presented before the conditioned stimulus because after the external unconditioned stimulus is removed the organism continues to respond to it as an internal stimulus (perception of pain) which occurs simultaneously with the presence of the conditioned stimulus. The perception of pain after the external stimulus is removed is of sufficient intensity to cause hand or finger withdrawal.

Before either the sensitization hypothesis or the duration-of-pain hypothesis can be ruled out, backward conditioning will have to be obtained without the use of a noxious stimulus.

#### REFERENCES

1. BARLOW, J. A. Secondary motivation through classical conditioning. *Amer. Psychologist*, 1952, **7**, 273.
2. ———. Secondary motivation through classical conditioning: A reconsideration of the nature of backward conditioning. *Psychol. Rev.*, 1956, **63**, 406-408.
3. BERITOV, I. S. Über die Individuell erworbene Tätigkeit des Zentralnervensystems bei Tauben. *Plug Arch. Physiol.*, 1926, **213**, 370-406.
4. ———. *Individually acquired activity of the central nervous system*. Tbilisi: GIZ, 1932.
5. BERNSTEIN, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *J. Gen. Psychol.*, 1934, **10**, 173-197.
6. BROGDEN, W. J. Animal studies of learning. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: Wiley, 1951. Pp. 568-612.
7. CARR, H. A., & FREEMAN, A. S. Time relationships in the formation of associations. *Psychol. Rev.*, 1919, **26**, 465-473.
8. CASON, H. Backward conditioning eyelid reactions. *J. Exper. Psychol.*, 1935, **18**, 599-611.
9. COPPOCK, H. W. Secondary reinforcing effect of a stimulus repeatedly presented after electric shock. *Amer. Psychologist*, 1951, **6**, 277.
10. DEESE, J. *The Psychology of Learning*. New York: McGraw-Hill, 1958.
11. FITZWATER, M. E., & REISMAN, M. N. Comparison of forward simultaneous, backward, and pseudoconditioning. *J. Exper. Psychol.*, 1952, **44**, 211-214.
12. FULTON, J. F. *Howell's Textbook of Physiology*. Philadelphia: Saunders, 1946.
13. GARDNER, E. *Fundamentals of Neurology*. Philadelphia: Saunders, 1952.
14. GOODSON, F. E., & BROWNSTEIN, A. Secondary reinforcing and motivating properties of stimuli contiguous with shock onset and termination. *J. Compar. & Physiol. Psychol.*, 1955, **48**, 381-386.
15. GREYER, W. F. Pseudoconditioning without paired stimulation encountered in attempted backward conditioning. *J. Compar. Psychol.*, 1937, **25**, 91-96.
16. HARRIS, J. P. Forward conditioning, backward conditioning, pseudoconditioning and adaptation to the conditioned stimulus. *J. Exper. Psychol.*, 1941, **28**, 491-502.
17. HILGARD, E. R. *Theories of Learning*. New York: Appleton, 1956.
18. KIMBLE, G. A. *Hilgard and Marquis' Conditioning and Learning*. New York: Appleton, 1961.

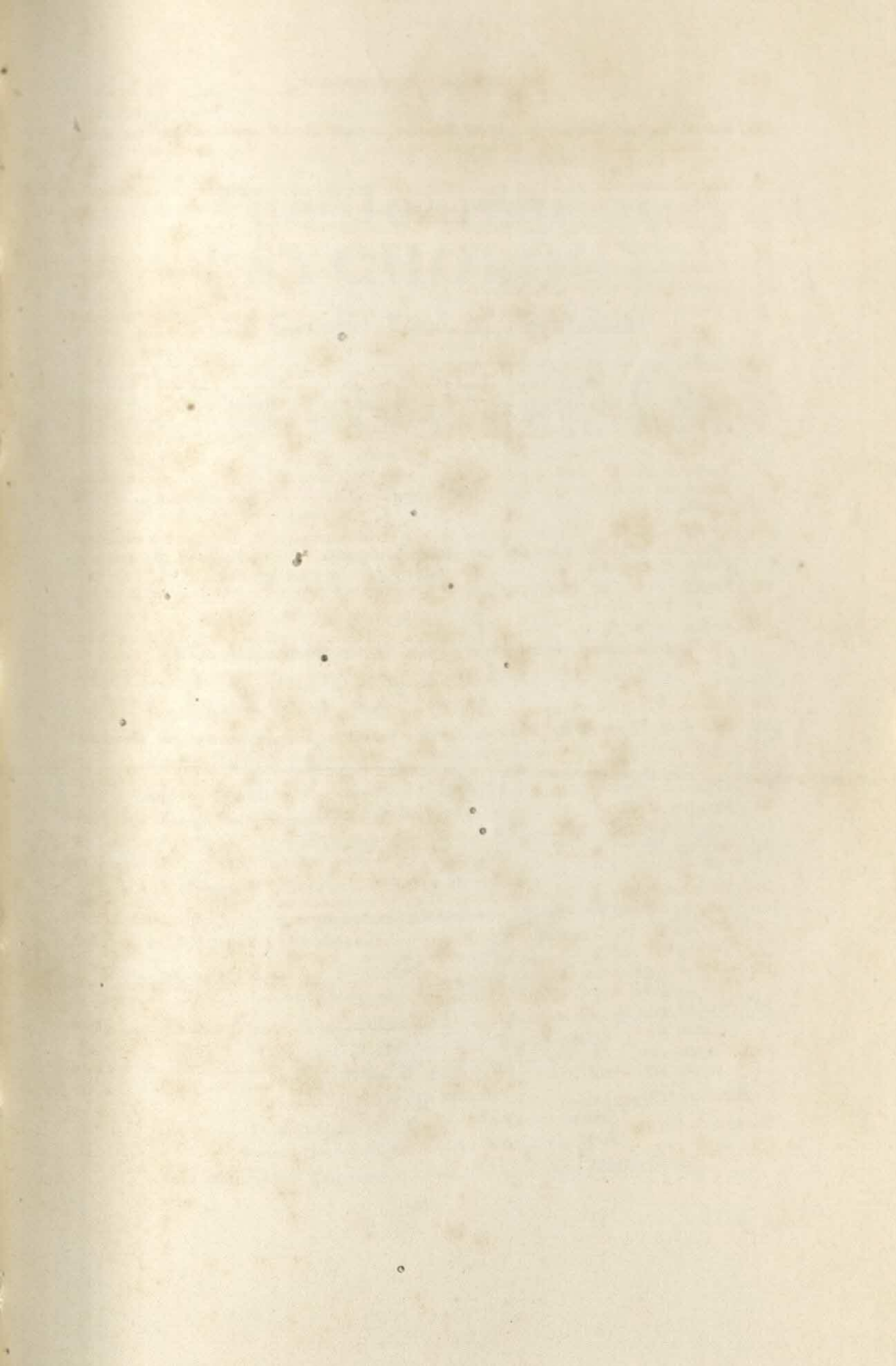


19. KREPS, E. M. The possibility of forming a conditioned reflex when the unconditioned precedes the indifferent stimulus. *Tr. Fiziol. Lab. Pavlova*, 1933, **5**, 5-20.
20. KRESTOVNIKOV, A. N. The essential conditions for the formation of conditioned reflexes. *Izv. Petogr. Nauchin In-ta. Lesgafta*, 1921, **3**, 197-240.
21. LAWSON, R. Learning and Behavior. New York: Macmillan, 1960.
22. NARBUTOVICH, I. O. The possibility of forming a conditioned reflex when the unconditioned defense stimulus precedes the indifferent stimulus. *Tr. Fiziol. Lab. Pavlova*, 1940, **9**, 371-396.
23. NEZHIDANOVA, Z. A. The formation of conditioned reflexes when the unconditioned stimulus precedes the indifferent stimulus. *Tr. Fiziol. Lab. Pavlova*, 1940, **9**, 353-359.
24. OSGOOD, C. E. Method and Theory in Experimental Psychology. New York: Oxford Univ. Press, 1953.
25. PAVLOV, I. P. Conditioned Reflexes. London: Oxford Univ. Press, 1927.
26. ———. Lectures on Conditioned Reflexes. New York: International Publishers, 1928.
27. PAVLOVA, A. M. Can a conditioned reflex be formed when the unconditioned stimulus precedes the indifferent stimulus. *Tr. Fiziol. Lab. Pavlova*, 1933, **5**, 21-31.
28. PETROVA, M. K. The prevalence of the inhibitory action of the unconditioned stimulus when it precedes the indifferent stimulus. *Tr. Fiziol. Lab. Pavlova*, 1933, **5**, 49-79.
29. PORTER, J. M., JR. Backward conditioning of the eyelid response. *J. Exper. Psychol.*, 1938, **23**, 403-410.
30. PRESSMAN, I. M. An attempt to form a conditioned reflex with children when the reinforcement precedes the signal. In A. C. Ivanov-Smolensky (Ed.), *Na puti k izuch vysshnykh form neirodinamiki kebënka*, Moscow: GIZ, 1934, 131-166.
31. RAZRAN, G. Backward conditioning. *Psychol. Bull.*, 1956, **53**, 55-69.
32. SMITH, M. P., & BUCHANAN, G. Acquisition of secondary reward as associated with shock reduction. *J. Exper. Psychol.*, 1954, **48**, 123-126.
33. SPOONER, A., & KELLOGG, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **60**, 321-334.
34. STROGANOV, V. V. The effects of the center of unconditioned stimuli upon that of conditioned stimulus. *Tr. Fiziol. Lab. Pavlova*, 1940, **9**, 360-370.
35. SWITZER, C. A. Backward conditioning of the lid reflex. *J. Exper. Psychol.*, 1930, **13**, 76-97.
36. VINOGRADOV, N. V. The emergence of new connections in the inhibition regions of the cortex. *Tr. Fiziol. Lab. Pavlova*, 1933, **5**, 33-48.
37. WOLFE, H. M. Time factors in conditioning finger withdrawal. *J. Gen. Psychol.*, 1930, **4**, 372-378.
38. ———. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. Gen. Psychol.*, 1932, **7**, 80-103.
39. WOODWORTH, R. S., & SHLOSBERG, H. Experimental Psychology. (Rev. ed.), New York: Holt, 1954.
40. YARBOROUGH, J. V. The influence of the time interval upon the rate of learning on the white rat. *Psychol. Monog.*, 1921, **30**, No. 2 (Whole No. 135).

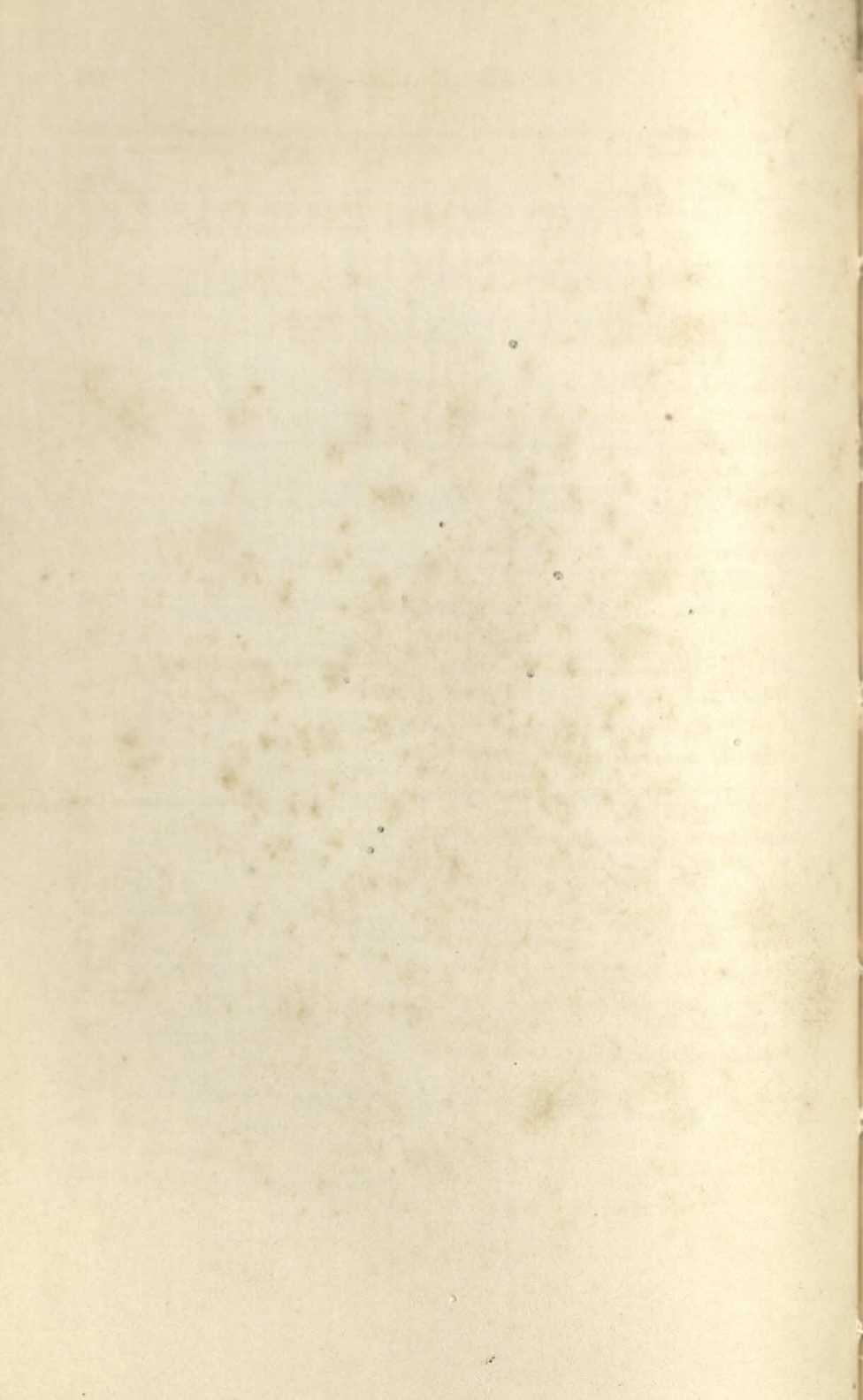
Psychology Department

Boston College

Chestnut Hill 67, Massachusetts







\$20.00 per annum  
\$15.00 per volume  
\$7.50 per half volume

QUARTERLY  
Two volumes per year  
Immediate publication

July, 1965  
Volume 60, Second Half

(Founded in 1935 by Carl Murchison)

# The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a liberal contract, and may not be resold for a period of five years from the date of publication.

JULY, 1965

(Manuscripts are printed in the order of final acceptance)

The effect of rate of presentation of paired associates on short-term retention . . .	147
BY DONALD L. HELMS AND B. L. KINTZ	
Attitudes toward colonialism, political dependence, and independence . . .	155
BY HARRISON G. GOUGH AND GIUSEPPE DI PALMA	
The semantic differential as a measure of subliminal message effects . . .	165
BY JERE R. HOAR AND EDWIN E. MEEK <sup>o</sup>	
A comparison of the organizational climates of Negro and white elementary schools . . .	171
BY HAROLD W. GENTRY AND JAMES B. KENNEY	
The relationship between expressed self-esteem and assumed similarity . . .	181
BY MAUREEN P. FITZGERALD	
Adaptive testing in an older population . . .	193
BY DENNIS I. GREENWOOD AND CHARLES TAYLOR	

(OVER)

Copyright © 1965, by The Journal Press  
Provincetown, Massachusetts, U. S. A.

All rights reserved

Published at 35 New St., Worcester, Mass.  
Second-class postage paid at Worcester, Mass.

Send all mail to 2 Commercial Street, Provincetown, Massachusetts



Contrast effects in extinction responding as a function of response effort during conditioning . . . . .	199
By THOMAS STACHNIK	
Identifying placebo reactors . . . . .	205
By HEBER C. SHARP	
The relationship of verbal ability and other cognitive variables to the open-closed cognitive dimension . . . . .	213
By SALVATORE V. ZAGONA AND LOUIS A. ZURCHER, JR.	
An empirical evaluation of the effect of item overlap on factorial stability . . . . .	221
By MILES S. ROGERS AND GERALD H. SHURE	
Delayed intellectual development in cerebral palsy children . . . . .	235
By SAMUEL HIRSCHENFANG AND JOSEPH G. BENTON	
Selective trace arousal . . . . .	239
By GERALD TOLCHIN AND JOHN CERASO	
Some correlates of inner directedness and other directedness . . . . .	245
By ANTHONY J. SMITH AND JOHN N. PARK	
Subjective colors: An objective-color artifact . . . . .	251
By B. J. FROST	
Recent advances in the psychology of masculinity and femininity . . . . .	255
By VYTAUTAS J. BIELIAUSKAS	
Speech avoidance among children with oral-communication defects . . . . .	265
By LOUIS LEREA AND BRUCE WARD	
The effect of previously learned verbal habits upon two response classes in verbal conditioning . . . . .	271
By PAULINE R. CLANCE AND T. R. DIXON	
S-R relationships and transfer of training . . . . .	277
By E. C. WORTZ AND A. C. MCTEE	
Logical thinking and paranoid schizophrenia . . . . .	283
By F. A. COYLE, JR. AND J. L. BERNARD	
The concept of cognitive dissonance . . . . .	291
By MAURICE N. RICHTER, JR.	
Visual-character detection . . . . .	295
By D. W. BROWN AND H. D. FRIEDMAN	
An investigation of coin-size estimates by orphaned children, normal children, and emotionally disturbed children . . . . .	301
By ANTHONY F. BRONZO, JR.	
Expectancy and validity . . . . .	309
By LEONARD W. FERGUSON	
ESP: Motivation as a factor in ability . . . . .	313
By RICHARD C. SPRINTHALL AND BARRY S. LUBETKIN	

## THE EFFECT OF RATE OF PRESENTATION OF PAIRED ASSOCIATES ON SHORT-TERM RETENTION\*<sup>1</sup>

*Department of Psychology, Ohio University*

DONALD L. HELMS<sup>2</sup> AND B. L. KINTZ

### A. INTRODUCTION

Several studies have been reported that demonstrate in short-term retention the influence of variables that are known to operate over longer intervals of time. Peterson, Saltzman, Hillner, and Land (9) studied short-term retention, using paired associates. They found improved recall with repetition, retroactive interference, and spontaneous recovery and suggested that short-term retention and long-term retention should be considered as basically similar until experimental evidence indicated otherwise. Keppel and Underwood (2) provide supporting evidence for that conclusion. Kintz and Zaffy (3), however, found that task difficulty had no effect upon short-term retention; but had a large effect upon long-term retention. They inferred from their results that two different memory mechanisms may be operative.

Of the many investigations in this area, those concerned with the rate of presentation are of particular relevance for the present study. Murdock (7) reported a significant difference in the recall of single paired associates between rates of one second and three seconds per pair. His comment that the results were in the expected direction—the faster rate produced poorer retention—is in agreement with a well-founded generalization in long-term retention.

But Peterson, Wampler, Kirkpatrick, and Saltzman (10), using rates of two seconds and four seconds, studied the effects of single, double-massed, and double-spaced presentations of paired associates on short-term retention and found data that showed the two-second rate to be superior to the four-second rate when one paired associate intervened between presentation and test; but when four intervening pairs were inserted between presentation and

\* Received in the Editorial Office, Provincetown, Massachusetts, on April 26, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was performed by the first author for a Master of Science thesis at Ohio University under the direction of the second author. The study was supported in part by the University Research Committee of Ohio University.

<sup>2</sup> Now at The University of Iowa.



test, the four-second rate showed better retention. Generally, these differences between rates held for all three types of presentation used.

A point in Peterson *et al.*'s study (10) that requires clarification is the distinction between the time intervals separating the presentation and the test (retention intervals) and the number of intervening paired associates in the intervals. Peterson *et al.* made the comparison between the two-second rate and the four-second rate on the basis of the number of intervening item pairs rather than in terms of the length of the time interval itself, although they plotted their data in terms of retention time. In another paper (8) Peterson more explicitly recognized this distinction, for further data supporting the previously discussed differences in rates were graphically presented with the number of intervening pairs as the independent variable.

The present study, by making use of numbers at the faster rates, was designed to eliminate the confounding of the number of intervening pairs and the length of the retention intervals, by equating rates on both variables. With this modification, differences between rates can be attributed more confidently to differences in exposure time and concomitant factors.

## B. METHOD

### 1. Subjects

The Ss were 60 male freshmen and sophomore students enrolled in introductory-psychology courses at Ohio University. Course points were given for voluntary participation in the experiment. Twenty Ss were randomly assigned to the presentation rates of one second, two seconds, and four seconds per pair.

### 2. Apparatus

Two Lafayette memory drums were used to present the paired associates. One drum served for the one-second rate, and the second drum presented both the two-second and four-second rates. The drums were placed on a table and inserted through an opening in a shield that separated S from E. Sliding covers on the drums permitted the exposure of different areas of the tape.

### 3. Materials

Paired associates were formed by pairing three-letter and four-letter single-syllable words. The words were randomly selected from Thorndike and Lorge's count of words occurring 10 or more times per million (11), after the elimination of homophones, contractions, proper nouns, and words denoting numbers.

A single list of paired associates was constructed. The list was composed of three equal-sized sublists in which each retention interval was tested three times or for a total of nine tests for the entire list. Within a sublist, the presentation and the testing of retention conditions overlapped. The assignment of pairs to retention conditions and the placement of conditions for each sublist were determined randomly. A particular word appeared only once in the list. To produce the proper spacing of retention intervals and to eliminate the possible advantage of initial position in the list, dummy pairs were inserted and used when necessary as filler items. These pairs were indistinguishable from pairs that were scored, but were never tested.

This basic list was used for all rates. However, in order to make the one-second and two-second rates more nearly equivalent to the four-second rate (in terms of time intervals between successive pairs), neutral material was added to the lists for these rates. The arrangement of the material in the three lists is given in Table 1. The neutral material consisted of the random grouping of the digits "2" through "9" into three-digit numbers. A pair of such numbers, per presentation, was used to fill the time intervals at the two-second rate; and a single three-digit number was used at the one-second rate. Thus, a time period comparable to one exposure at the four-second rate was filled with one paired associate plus one neutral exposure at the two-second rate and one paired associate plus three neutral exposures at the one-second rate. The introduction of this neutral material was designed to reduce rehearsing which might have occurred if the time interval was unfilled.

The basic list was modified, so that the presentation of any given pair was in exactly the same position in all lists—in terms of the time interval from the beginning of the list and time intervals to any other presentation. The total time of presentation was held constant for all lists.

#### 4. Procedure

The *S* was ushered into the experimental room and seated facing the memory drum. Then *E* took his position behind the screen and read standard verbal-learning instructions to the *S*. The *S* was told to say aloud everything that appeared in the aperture of the drum and to supply the missing response by saying it aloud when the stimulus word appeared alone. Guessing was encouraged. No questions were answered by *E* once the drum was started.

The method of presentation employed was a modified version of that used by Peterson *et al.* (9), Kintz (3), and Kintz and Zaffy (4). In place of the usual anticipation technique, in which successive turns of the drum expose



TABLE 1  
SAMPLE ITEMS FROM THE THREE LISTS\*

Four seconds	Two seconds	One second
MUSE—JEAN	MUSE—JEAN	MUSE—JEAN
	693—463	693 463 556
BUZZ—FACT	BUZZ—FACT	BUZZ—FACT
	BUZZ—	BUZZ—
		397 523 437 933 766 689
BUZZ—	397—523	
	437—933	
LEND—WREN	LEND—WREN	LEND—WREN
	944—722	944 722 227
PEAK—FOLD	PEAK—FOLD	PEAK—FOLD
	629—737	629 737 588
REEL—SAT	REEL—SAT	REEL—SAT
	756—226	756 226 435
FEW—MOSS	FEW—MOSS	FEW—MOSS
	REEL—	REEL—
		837 789 822
REEL—	822—837	
	LEND—	LEND—
		325 835 296
LEND—	296—325	
	PEAK—	PEAK—
		983 859 272
PEAK—	272—983	

\* In this sample the retention interval for BUZZ— is 0 seconds, for REEL— four seconds, and for LEND— and PEAK— 16 seconds, regardless of the presentation rate.

the stimulus alone followed by the S-R pair, a single exposure of the S-R pair was used. Testing was accomplished by presenting the stimulus alone after zero, four, or 16 seconds had elapsed. Each S was tested nine times at each retention interval at one of the three rate conditions. The drum revolved continuously to present four consecutive repetitions on the same list.

The length of the memory tape was equal to the length of the list for the four-second rate; so successive repetitions were given by each complete revo-

lution of the drum. The lists for the two-second and one-second rates were two times and four times the length of the four-second list respectively, and this required *E* to move sliding drum covers to expose successive portions of these lists.

### C. RESULTS

The nine retention tests for each *S* at each retention interval were assigned scores of one for a correct response and zero for an incorrect response. The mean proportions of responses correctly made are presented in Table 2.

TABLE 2  
MEAN PROPORTIONS OF ITEMS CORRECTLY RECALLED

Retention interval and presentation number	Presentation duration		
	One second	Two seconds	Four seconds
Zero seconds			
1	.93	.99	.99
2	.98	.99	.99
3	.99	1.00	1.00
4	.99	1.00	.99
Four seconds			
1	.31	.34	.57
2	.39	.43	.71
3	.49	.48	.83
4	.59	.63	.87
Sixteen seconds			
1	.09	.10	.29
2	.20	.27	.52
3	.30	.39	.67
4	.42	.46	.78

An analysis of variance of the data was performed using a mixed design. Due to heterogeneity of both form and variance among the treatment groups, and in view of other restrictive assumptions often unsatisfied in "mixed" designs (designs in which some dimensions have repeated measures on the same *Ss*), all tests of significance were required to exceed the tabled values at the .025 level (5, p. 83).

The analysis of variance of the total number of correct responses for all conditions is summarized in Table 3. All *F*s are significant, with the exception of the rate-by-repetitions interaction.

The differences between rates at the four-second and 16-second retention intervals did not appear at the zero-second interval because of the nearly perfect performance of all *Ss* on the immediate-recall measure. The significant



TABLE 3  
SUMMARY OF THE ANALYSIS OF VARIANCE OF THE TOTAL NUMBER OF  
CORRECT RESPONSES

Source	<i>df</i>	<i>MS</i>	<i>F</i>
Between Ss	59		
Rates (R)	2	248.651	33.81*
Error (between Ss)	57	7.354	
Within Ss	660		
Retention Intervals (I)	2	1945.955	421.20*
Repetitions (T)	3	148.641	171.05*
Interaction			
Intervals-by-repetitions	6	33.093	43.54*
Intervals-by-rates	4	55.983	12.12*
Rates-by-repetitions	6	1.383	1.59
Intervals-by-rates-by-repetitions	12	1.631	2.15*
Error (within)	627		
Error (I and I-by-R)	114	4.620	
Error (T and R-by-T)	171	.869	
Error (I-by-T and I-by-R-by-T)	342	.760	
Total	719		

\*  $p < .025$ .

rate-by-retention-interval interaction reported in Table 3 reflects the even more homogeneous means at this interval when the repetitions are averaged.

The repetitions-by-retention-interval interaction is primarily a function of the lack of spread at the zero-second retention interval.

The differences between retention-interval means are highly significant. Table 2 indicates the magnitude of the decrease in retention over time for each presentation rate. There is also an increase in the proportion of correct recall as the number of repetitions increases.

Significant differences between the rate means were obtained [see Table 3], but the comparisons of major interest are those between the rates for the first presentation. A partial mixed-design analysis of variance shows the differences between the rates of the first-presentation data to be significant beyond the .001 level. Appropriate  $t$  tests show the means of the one-second and two-second rates not to differ significantly ( $t = 1.14$ ,  $df = 57$ ). Significant differences were found between rates of one second and four seconds ( $t = 6.20$ ,  $df = 57$ ) and rates of two seconds and four seconds ( $t = 5.07$ ,  $df = 57$ ).

#### D. DISCUSSION

The striking differences among the retention intervals are in general agreement with a substantial body of evidence in short-term retention (6). The nearly perfect recall at the zero-second interval is a commonly reported out-

come [see Peterson *et al.* (9)]. Some authors have argued that a test at this interval should be regarded as an index of the accuracy of perception rather than as a measure of retention. In any case, the zero-second condition in the present study served primarily to insure that retention losses at longer intervals would not be overestimated; proportions correctly recalled at the four-second and 16-second retention conditions are meaningful only if recall is very close to 1.00 at zero seconds [but see Underwood (12) for a discussion of related problems].

One point of departure from many studies of short-term retention is the sharp drop in proportions correct from the zero-second to the four-second retention test. The data of Table 2 illustrate this decrement. Melton's discussion and summary of pertinent data (6) revealed a much smaller drop over comparable time periods in other studies. However, Melton discussed short-term-retention experiments in which the retention intervals contained only material designed to prevent rehearsing, in contrast to the overlapping presentation and testing conditions in the present study. The experiment by Peterson *et al.* (10), which is basically similar to this study, also showed large retention losses over periods of four seconds and longer.

The learning that is represented by improvement over repetitions, while certainly a familiar phenomenon, is difficult to relate directly to other investigations of short-term retention. The efficacy of repetitions of items has been demonstrated [Hellyer (1), Peterson *et al.* (9), and Waugh (13)] in a variety of situations, but no previous studies of short-term retention have reported repetitions of a list in the manner used here.

The variable of central importance to this experiment is the relationship between the presentation duration and the retention interval. In agreement with findings by Peterson *et al.* (10), an interaction did occur between these two factors; however, the interaction was interpreted as due primarily to nearly equal means at the zero-second retention interval. Peterson *et al.* further reported that the two-second rate is superior to the four-second rate when one paired associate intervened between presentations and tests or at retention intervals of two seconds and four seconds respectively. The four-second recall data in Table 2 show no indication of such an effect. Furthermore, retention losses for all rates are roughly parallel, in contrast to Peterson *et al.*'s findings (10). They found the superiority of the two-second rate over the four-second rate at very brief intervals to be reversed over longer intervals, with the two-second rate very inferior to the four-second rate at recall intervals of eight seconds and longer.



## E. SUMMARY

Each of three groups of 20 Ss were presented with a list of paired associates at rates of one second, two seconds, and four seconds. Within each rate the Ss were tested at retention intervals of zero, four, and 16 seconds. Four presentations of each list were given. Previous work had suggested that a two-second rate produces better recall than a four-second rate at retention intervals of four seconds or less, but falls sharply below a four-second rate when the retention period is eight seconds or longer. These results were *not* duplicated in the present study; so the authors concluded that the length of the retention interval, in addition to the interfering material in the interval, must be taken into account in the study of retention.

## REFERENCES

1. HELLYER, S. Supplementary report: Frequency of stimulus presentation and short-term decrement in recall. *J. Exper. Psychol.*, 1962, **64**, 650.
2. KEPPEL, G., & UNDERWOOD, B. J. Proactive inhibition in short-term retention of single items. *J. Verb. Learn. & Verb. Behav.*, 1962, **1**, 153-161.
3. KINTZ, B. L. Short-term retention and long-term retention as a function of practice. *J. of Psychol.*, 1965, **59**, 309-314.
4. KINTZ, B. L., & ZAFFY, D. J. Short-term and long-term retention and task difficulty. *J. of Psychol.*, 1965, **59**, 229-232.
5. LINDQUIST, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton, 1953.
6. MELTON, A. W. Implications of short-term memory for a general theory of memory. *J. Verb. Learn. & Verb. Behav.*, 1963, **2**, 1-21.
7. MURDOCK, B. B., JR. Direction of recall in short-term memory. *J. Verb. Learn. & Verb. Behav.*, 1962, **1**, 119-124.
8. PETERSON, L. R. Associate memory over brief intervals of time. *J. Verb. Learn. & Verb. Behav.*, 1963, **2**, 102-106.
9. PETERSON, L. R., SALTZMAN, D., HILLNER, D., & LAND, V. Recency and frequency in paired-associate learning. *J. Exper. Psychol.*, 1962, **63**, 396-403.
10. PETERSON, L. R., WAMPLER, R., KIRKPATRICK, M., & SALTZMAN, D. Effect of spacing presentations on retention of a paired associate over short intervals. *J. Exper. Psychol.*, 1963, **66**, 206-209.
11. THORNDIKE, E. L., & LORGE, I. The Teacher's Word Book of 30,000 Words. New York: Teachers College, Columbia University, 1944.
12. UNDERWOOD, B. J. Degree of learning and the measurement of forgetting. *J. Verb. Learn. & Verb. Behav.*, 1964, **3**, 112-129.
13. WAUGH, N. C. The effect of intralist repetition on free recall. *J. Verb. Learn. & Verb. Behav.*, 1962, **1**, 95-99.

Department of Psychology  
Ohio University  
Athens, Ohio 45701

## ATTITUDES TOWARD COLONIALISM, POLITICAL DEPENDENCE, AND INDEPENDENCE\*

*Institute of Personality Assessment and Research and Department of Political Science,  
University of California, Berkeley*

HARRISON G. GOUGH AND GIUSEPPE DI PALMA

### A. INTRODUCTION

The movement of former colonies and dependent states toward political independence and membership in the community of nations is clearly one of the significant political phenomena of today. In fact, the trend is so visible and tangible that the conceptual phrase "the emerging nations" is in almost everyday currency.

Whenever political or social events begin to take shape and to move in a particular direction it is to be expected that attitudes with respect to these changes will crystallize. From the standpoint of attitude and opinion research two problems present themselves: first, to develop a brief and reliable index of the attitudinal dimension (i.e., favorable or unfavorable sentiments concerning the trend); and second, to relate this attitudinal continuum to sociological, demographic, and personological factors with which it covaries and which therefore may assist in its clarification.

The initial requirement, it follows, is to develop an adequate measure of attitudes toward colonialism and political independence, and it is this task to which this paper is addressed.

### B. DEVELOPMENT OF THE SCALE

Our first step was the writing of a set of items that embodied a wide range of opinions concerning political independence and dependence. An attempt was made to avoid extremes of negativism or positivism, so as to reduce the influence of desirability and acquiescence artifacts; and an attempt was made to express all views in a nonpejorative manner. A procedure to be described later was also adopted so as to overcome any biases in expression that the authors were not able to hold in check by conscious effort.

One hundred items were written, of which 34 were scored for anticolonialism on the basis of apriori judgment; these 34 items were those of unambig-

\* Received in the Editorial Office, Provincetown, Massachusetts, on May 6, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



uous meaning, for which a clear designation of procolonialism or anticolonialism could be made. Eight representative items are listed below (with scoring for anticolonialism):

1. All countries still ruled by colonial powers should immediately be granted unconditional and total independence (agree).
2. When a colonial power grants independence to one of its colonies, that country should not be allowed to maintain any political or economic control over the former colony (agree).
3. The best guarantee for stability and progress of a former colony is strict cooperation with the country which formerly ruled it (disagree).
4. A difficulty with granting independence too soon to a colony is that the new country will face economic and social problems that it cannot handle by itself (disagree).
5. Colonial powers should grant independence to any of their subject countries that expressly request this (agree).
6. The white nations should still bear the main burden of leading the colored nations toward progress (disagree).
7. A colony that is not granted independence has the right to achieve it by rebellion (agree).
8. Nowadays, colonial control of a backward country by a more advanced country usually improves the social welfare of the colony (disagree).

The complete 100-item schedule was given to 52 students in a social psychology class at the University of California, with responses of "agree," "uncertain," and "disagree" being requested.<sup>1</sup> Scoring was on the basis of the 34-item key just mentioned. A mean of 16.4 was obtained, with a range of eight to 26. Next, each of the 100 items was correlated with the 34-item score to identify new items closely aligned with the original dimension and to identify any psychometrically inconsistent items among the original 34.

Thirty-five items were selected in this analysis, including 20 new items and excluding 19 of the original 34. This shift was intentional and was undertaken in the hope of broadening the basis of assessment by bringing in attitudinal material of a less-obvious or less-emphatic nature, provided that these new items would be fully consonant with the original dimension. An example of these new items is, "In politics what counts is not so much the goal as the effectiveness of the means selected for achieving that goal." The statistical analysis permitted the scoring of "uncertain" responses in 23 of the 35 items.

A consideration in selecting the 35 items was the control of the desirability of options. Some items were included on which a high percentage of students gave a scored response, and an equal number on which a low percentage ob-

<sup>1</sup> We thank Professor E. Ballachey for permitting this testing.

tained a point. An effort was made to include a number of items with a 50 per cent incidence of scored answers. This "centering" was apparently successful because the mean of the 52 protocols rescored for the 35-item key was 18.2, and the range was five to 27.

The third step in the analysis was to compute item-*vs.*-total-score correlations for each of the 35 items (using the new total score) and Guttman reproducibility coefficients (8) for each item. Twenty-eight of the items satisfied the minimal requirements of these two criteria and were retained in a prefinal form of the attitude scale. At this point, the 28-item scale contained items (a) having clear psychological and psychometric relevance to the anticolonialism continuum, (b) equally distributed over pro and con positions, and (c) possessing properties of unidimensionality and reproducibility.

The fourth step in the development of the measure was to apply the 28-item version in a new sample of 120 students.<sup>2</sup> On each item unit weights were assigned for the anticolonial response or responses, and these weights were correlated with total score. From this analysis the 18 best items were retained, with scoring and item-*vs.*-total-score point-biserial correlations as reported in Table 1.

For the class of 120 students, the 18-item scale yielded a mean of 7.96, a standard deviation of 3.13, and a range of one to 18. There was a slight (and statistically insignificant) sex difference, with females scoring higher. On the first sample of 52 students, the 18-item scale gave a mean of 8.88, a standard deviation of 3.46, and a range of two to 16. In the sample of 52, the correlations of the final scale with the earlier scales were .74 (with the first 34-item scale), .94 (with the second 35-item scale), and .96 (with the third 28-item scale).

In the sample of 120 students used to reduce the 28-item scale to the final 18-item version, the split-half corrected reliability coefficient was .67. On a sample of 88 students (described later) the corrected split-half coefficient was +.71.

### C. COMPARISON WITH OTHER VARIABLES

Perhaps the best known and most widely used attitude scale developed to date is the F Scale for Authoritarianism (1). Any new scale, it would seem, should be checked against the F Scale to make sure that the new measure is not just an indirect (or inefficient) index of the F Factor. Some writers [e.g. (2, 3)] have suggested that response sets, such as desirability and acquiescence,

<sup>2</sup> Professor Ballachey was kind enough to permit this testing in a subsequent class in social psychology.



TABLE 1  
ITEMS RETAINED IN THE FINAL FORM OF THE ANTICOLONIALISM ATTITUDE SCALE, WITH  
SCORING, PERCENTAGE OF ANTICOLONIAL ANSWERS, AND CORRELATION  
OF ITEMS WITH TOTAL SCORE

No.	Item and scoring	Per cent correct	Correlation with total score
1.	Any cultural, linguistic, or ethnic group which wishes to withdraw from a state and set up a new state should have the right to do so (agree, uncertain).	30.8	.42
2.	Colonial powers should grant independence to any of their subject countries that expressly request this (agree).	51.7	.40
3.	No matter how underdeveloped, a colony that is granted real independence will in a reasonable period of time be able to resolve its own internal problems without external intervention (agree, uncertain).	47.5	.32
4.	Experience after World War II has shown that the colonial territories which achieved the status of independent nations (e.g., Tunisia) can and do act with responsibility and good sense in matters of peace and international understanding (agree, uncertain).	63.3	.26
5.	A colony that is not granted independence has the right to achieve it by rebellion (agree, uncertain).	67.5	.29
6.	There is little to lose and something to gain if former colonies maintain some ties with their former colonial powers (uncertain, disagree).	18.3	.25
7.	If necessary, the United Nations and individual countries should impose sanctions against those countries which do not grant independence to their colonies (agree, uncertain).	40.0	.47
8.	The main international task of the newly developing nations should be that of creating a third international force, independent of the pressures from both Russia and the West (agree, uncertain).	42.5	.33
9.	Some of the colonial powers tend to grant independence to their colonies only to establish more subtle, informal systems of control over them (agree, uncertain).	65.0	.26
10.	The case of Algeria has shown that national liberation movements do not result from small elite groups working in a context of general apathy, but are on the contrary large mass movements appealing to the basic needs of the local peoples and commanding their support (agree, uncertain).	61.7	.34

TABLE 1 (*continued*)

No.	Item and scoring	Per cent correct	Correlation with total score
11.	With help from international cooperation and with freedom from any form of latent exploitation, the newly developing nations can employ the great technological resources of today to accomplish in a few years what more developed nations have accomplished in the last 100 years (agree).	67.5	.27
12.	Nowadays, colonial control of a backward country by a more advanced country usually improves the social welfare of the colony (disagree).	30.0	.47
13.	The general impact of colonialism in the past, e.g., Britain's control of India, Belgium's control of the Congo, etc., has been deleterious to the cause of human freedom, socio-economic progress, and understanding among peoples and races (agree).	36.7	.23
14.	Substantially, if not formally, a large part of Latin America is in a position comparable to that of a colony of the United States (agree, uncertain).	40.8	.28
15.	If necessary for the maintenance of international equilibrium, we should occupy the territories of strategic importance that otherwise are likely to fall into the hands of potential enemies (uncertain, disagree).	45.0	.47
16.	No world power should withdraw from a colony if that colony is of strategic importance to the maintenance of the international balance of power (disagree).	45.8	.28
17.	The actions of the United States toward Cuba are comparable to those of a colonial power that tries to suppress movements of national liberation within its colonies (agree, uncertain).	25.8	.41
18.	In politics what counts is not so much the goal as the effectiveness of the means selected for achieving that goal (uncertain, disagree).	45.8	.37

play a significant role in determining scores on the F Scale. To the extent that the foregoing is true, one would have an additional reason for wishing to demonstrate independence between the new measure and the F Scale. Further, intelligence may play a role in the genesis of procolonial versus anticolonial attitudes; but one would hope that it would not be the main determiner of scores on the scale. In short, it is important in developing an attitude measure to see if it is free of relationships that could bias its meanings.

To check on the correlations with the F Scale, intellectual ability, and



prior scales of political opinion, the anticolonialism measure—along with the F Scale, McClosky's scales for anomie and conservatism (9), Rokeach's scales for dogmatism and opinionation (10), and the College Vocabulary Test (7)—was given to a sample of 88 University of California students.

For this sample of 88 students, the mean score on the anticolonialism scale was 7.84, and the standard deviation was 2.89. Its correlation with the F Scale was  $-.20$ . This value, in the expected direction (less authoritarianism associated with greater opposition to colonialism), is not high enough to suggest that one scale can be substituted for the other. With McClosky's scales, the coefficients were  $.16$  with anomie and  $-.17$  with conservatism; again in the expected direction, but encouragingly low.

The correlation with Rokeach's scale for dogmatism was  $.05$ , and with left opinionation the correlation was  $.56$ . Anticolonialism, one may conclude, has a definite relationship to the more-generalized political attitudes assessed by Rokeach's measure. Paralleling the foregoing is a  $-.33$  correlation with right opinionation. With total opinionation (an index of one's tendency to adopt categorical and unmodulated political postures), the anticolonialism scale correlates  $.14$ . Finally, with the vocabulary test, the correlation is an insignificant  $-.04$ .

In a sample of 46 college females, the anticolonialism scale was correlated with Dicken's scales for social desirability and acquiescence on the California Psychological Inventory (4, 5). In this sample, anticolonialism yielded a mean of 7.61 and a standard deviation of 2.23. Its correlation with social desirability was  $.02$ ; with acquiescence,  $.05$ . These values indicate that the efforts made to keep the anticolonialism scale free from response-set weaknesses have been moderately successful.

#### D. CROSSCULTURAL APPLICATION

The anticolonialism scale was developed with the explicit hope of being applicable in research on political and social attitudes anywhere: i.e., that it would function validly in translation and crosscultural usage. To secure preliminary evidence on this issue, the 18-item instrument was translated into French,<sup>3</sup> and arrangements were made for the testing of a bilingual sample with both English and French editions.<sup>4</sup>

<sup>3</sup> The translation was prepared by H. C. de Bettignies.

<sup>4</sup> We thank Professor R. E. LaPointe of the Université de Montréal, who conducted the testing with the French and English editions. A copy of the French version as used in this testing has been placed on file in the American Documentation Institute, Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D.C. 20540. Order Document No. 8416, remitting \$1.25 for photoduplicates or \$1.25 for microfilm.

Ninety-five students at the University of Montreal, Quebec, Canada, took the French version; 90, the English. Of these, 30 completed both forms. For the French version the mean score was 10.60, and the standard deviation was 2.41. On the English version the mean was 11.10; the standard deviation, 2.74. These values are fairly similar and imply that the two editions function in an equivalent way. For the 30 students who took both versions, a correlation of .72 was obtained between the scores on the French and the English tests. Viewed as an equivalent-form reliability coefficient, this correlation is of the magnitude typically observed for measures of attitudes, personality, and interests.

One should note that the means on both the French and the English versions are higher than the means on the English version obtained on the samples tested in California. Whether these results are interpretable as something other than psychometric variability cannot be determined on the basis of the data available.

An Italian version<sup>5</sup> of the scale was prepared and was given to a sample of 34 students at the University of Bologna, Italy.<sup>6</sup> For these students the mean score was 11.18; the standard deviation, 2.86.

#### E. PSYCHOLOGICAL IMPLICATIONS

As a final source of information on the anticolonialism scale, scores on the measure were correlated with adjectival descriptions made of each other by a sample of 45 girls. For this purpose, each girl completed an Adjective Check List (6) on five of her friends, so that for each girl a panel of five observers was defined. By counting the number of times an adjective had been checked and by giving two points whenever an adjective was designated as "very highly characteristic," a score was derived for each word. These scores could range from zero to 10.

In a search for those words that identify girls high on anticolonialism and girls ranking low, the anticolonialism scores for the 45 girls were correlated with each of the 300 adjectival scores. The purpose of the analysis was to delineate a psychological portrait of the subject with high scores and of the subject with low scores. Each of the 45 girls was a member of a living group; so the descriptions may be presumed to have some degree of interpersonal validity.

<sup>5</sup> A copy of the Italian version of the scale has been deposited with the American Documentation Institute. Order Document No. 8416, remitting \$1.25 for photoduplicates or \$1.25 for microfilm.

<sup>6</sup> This testing was arranged through the courtesy of Professor Mario Battacchi of the University of Bologna.



A great many individual words had correlations of a statistically significant magnitude, but any attempt to list all of them would defeat the purpose of the analysis, which is heuristic—not definitional. A choice will be made to emphasize themes and patterns in the relationships. Two words of direct relevance to the intentions of the scale had significantly positive correlations: independent and progressive. Girls with high scores on anticolonialism were described by their peers as independent and progressive. They were also described as adventurous, frank, handsome, attractive, and good-looking. The syndrome of associated personological characteristics appears to be one stressing initiative, verve, independence, and self-confidence.

Among the words used to describe girls with the lowest scores on anticolonialism, the most significant correlation ( $-.43$ ) was with "slow." Inhibited, intolerant, stolid, and planful were also used to describe girls with low scores. The patterns may seem to favor the high scorer and, indeed, there may have been a slight trend in this direction; however, to balance the picture negative qualities (such as carelessness and impulsivity) were observed among the high scorers. And positive qualities, such as foresight, good judgment, and loyalty were mentioned in describing the low scorers.

The constellation of favorable attributes characterizing the girls with higher scores may stem from the fact that these girls reflect in their attitudes the progressive and liberal viewpoints which are ascendant on today's American campus. The same syndrome might not prevail among individuals whose anticolonialism finds little environmental support. The obvious limitation imposed by the lack of adjectival data for males should also be noted.

## F. DISCUSSION

The anticolonialism scale described in this report appears to be a valid and useful tool for the study of political and social attitudes. New nations are coming into existence almost daily, and the phenomenon of the growing independence of former colonies and dependent states is a significant political occurrence of our time. Attitudes toward colonialism, and toward the hastening or retarding of the time of its disappearance are proper current subjects of study.

To make such study possible, one needs a brief, valid, and reliable measure of anticolonial attitudes—one that is free of nuisance relationships to response-set artifacts and to measures of other political attitudes. Although not essential perhaps, one hopes for a measure that is valid in translation and in cross-cultural usage.

The anticolonialism scale offered in this paper seems to possess the quali-

ties just enumerated. It correlates little if at all with acquiescence or desirability; and it is only slightly related to the California F Scale or to measures of dogmatism, general opinionation, anomie, and conservatism. It is also free of any sizeable correlation with verbal intelligence. At the same time, its content validity is adequate; and its personological implications (insofar as they have been determined) are in line with theoretical expectations.

## REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The Authoritarian Personality*. New York: Harper, 1950.
2. BASS, B. M. Authoritarianism or acquiescence? *J. Abn. & Soc. Psychol.*, 1955, **51**, 616-623.
3. CRONBACH, L. J. *Essentials of Psychological Testing* (Second ed.). New York: Harper, 1960.
4. DICKEN, C. F. Good impression, social desirability, and acquiescence as suppressor variables. *Educ. & Psychol. Meas.*, 1963, **23**, 699-720.
5. GOUGH, H. G. *Manual for the California Psychological Inventory* (Rev. ed.). Palo Alto, California: Consulting Psychologists Press, 1964.
6. GOUGH, H. G., & HEILBRUN, A. B., JR. *Adjective Check List Manual*. Palo Alto, California: Consulting Psychologists Press, 1965.
7. GOUGH, H. G., & SAMPSON, H. *The College Vocabulary Test*. Berkeley, California: the University of California Institute of Personality Assessment and Research, 1954.
8. GUTTMAN, L. A basis for scaling qualitative data. *Amer. Sociol. Rev.*, 1944, **9**, 139-150.
9. MCCLOSKEY, H. Conservatism and personality. *Amer. Polit. Sci. Rev.*, 1958, **52**, 27-45.
10. ROKEACH, M. *The Open and Closed Mind*. New York: Basic Bks., 1960.

*Institute of Personality Assessment and Research*  
*University of California*  
*2240 Piedmont Avenue*  
*Berkeley, California 94720*





## THE SEMANTIC DIFFERENTIAL AS A MEASURE OF SUBLIMINAL MESSAGE EFFECTS\*<sup>1</sup>

*Department of Journalism, University of Mississippi*

JERE R. HOAR AND EDWIN E. MEEK

### A. INTRODUCTION

Since 1863 the use of visual and auditory subliminal exposure has been reported. To summarize, the data seem to indicate (a) that messages can be received below the threshold of conscious awareness (4), (b) that discrimination may take place at subthreshold intensities (2), (c) that subjects can make discriminatory responses even when they are not able to report the stimulus correctly (5), (d) that influence on behavior other than discrimination behavior can sometimes be achieved (1); but (e) that persuasion—i.e., getting subjects to take prescribed actions—does not take place under conditions and with techniques presently in use in controlled experiments (3).

From results obtained by Bach and Klein (1) and others, the present authors theorized that subliminal messages of a persuasive nature that are communicated (i.e., that reach the nervous systems of subjects) have effects whether or not the magnitude of the attitude changes brought about are great enough to affect overt behavior. Further, they theorized that this attitude change can be demonstrated with the proper instruments.

Physical demonstration of the foregoing theories might imply that, when subliminal messages fail to persuade, failure may be due to stimulus competition, lack of subject motivation, or to a stimulus of short duration or of low intensity.

### B. METHOD

Thirty-three Ss, subdivided into two experimental groups and one control group, took part in the study. All Ss were journalism students, and all but one were undergraduates.

In one part of the experiment, both experimental groups were urged subliminally to "vote Wehle" and "eat Yble." In another part, the experimental

\* Received in the Editorial Office, Provincetown, Massachusetts, on May 11, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This study was supported in part by a grant from the Committee on Faculty Research, University of Mississippi.



groups were "exposed" to each of the following filmed subliminal messages: Free Press Passive, Press Room Calm, Party Leader Strong, Lobbyist Valuable, Legislature Fair, and Reporter Bad.

All messages were projected during the showing of an Ohio State University film entitled *Legislative Reporter*. The showing lasted approximately 20 minutes and, during the showing, each subliminal message was projected 12 times.

Subliminal messages were projected onto the movie screen by means of a standard 35-roll film projector. The projector was placed about one-half inch from the focal-plane shutter in a 4 × 5 Speed Graphic camera. Ground glass and lens had been removed from the camera leaving only the shell and the back shutter.

Both camera and projector were mounted on a base attached to a standard tripod. Subliminal messages were projected from a darkened projection room, the subliminal apparatus being about two feet from the movie projector. The projection room was soundproof, and double exposure of the messages (while the experimenter cocked the focal-plane shutter) was avoided by means of a hinged metal flap that could be lowered over the lens opening.

An advantage of the focal-plane shutter was that many exposures could be tested to determine the average limen. Experiments were tried using speeds ranging from one-fiftieth to one-thousandth of a second. It seemed that a speed of one four-hundredth of a second was the speed nearest the threshold of awareness of nearly all persons tested.<sup>2</sup> Careful attention was given, so that messages were not exposed during periods in which the film was projected at low intensity (i.e., fade-ins and fade-outs). During the testing, Ss seemed not to be aware of the messages; however, one woman reported seeing letters on one occasion and wondered if there were a flaw in the film.

All Ss were pretested from three days to one week before "exposure" to the subliminal influences and were given an identical test as soon as they had seen the movie.

The control group saw the movie with no messages. One experimental group saw the movie and was exposed to eight subliminal messages, but was not warned that any clues were to be given. The second experimental group saw the movie and all eight messages, but was told to watch with particular care for a message of importance that would be presented.

Because the film used is a journalism-training film, subjects were told that

<sup>2</sup> The authors arrived independently at the same exposure speed and method of projection as used by Parker (7).

the film showing substituted for one of the regular class lectures on the subject "legislative reporting." The nature of the film is such that it could have applicability in different classes.

Two separate tests were given on each test date. In one, respondents were to choose from a ballot and vote for a candidate for office. They were to assume that they knew nothing about any of the candidates, who were listed as Elhew, Wehle, and Hewle. They also were to choose a breakfast food from a shelf of unfamiliar brands: Orser, Ybel, and Zern. The order in which the foils were presented on the ballot was rotated.

The second test was a semantic-differential test in which six concepts were judged in terms of each of the following 10 dichotomies: good—bad, worthless—valuable, fair—unfair, unpleasant—pleasant, weak—strong, heavy—light, small—large, active—passive, slow—fast, and agitated—calm. Scales were chosen on an *a priori* basis from among those tested by Osgood, Tannenbaum and Suci (6). Factor titles were placed on the pages in a random order, and a single page was used for each concept.

Concepts tested before and after exposure to subliminal messages were as follows: Party Leader, Lobbyist, Legislature, Reporter, Press Room, and Free Press. Subliminally, between the two tests, subjects were "exposed" to each of the following messages: Vote Wehle, Eat Ybel, Free Press Passive, Press Room Calm, Party Leader Strong, Lobbyist Valuable, Legislature Fair, Reporter Bad.

Some of the subliminal suggestions were in accord with the objective message of the film (Lobbyist Valuable); some were opposed (Press Room Calm); others were neither in accord nor opposed (Legislature Fair).

### C. RESULTS

Table 1 shows the changes in selections that took place after subjects had seen the messages "vote Wehle and eat Ybel." In three of four instances, the experimental groups moved in the directions urged. Control-group choices universally moved in the opposite directions. However, changes are not significant ( $p = .05$ ).

Results obtained from the semantic differential are as follows. Eight of 12 averaged scores for the experimental groups represent changes in the directions urged. For the control group, three of the changes are negative and three are positive.

None of the changes is significant at the .05 level. In computing chi square, subjects were combined into two groups: those voting in the motivated direc-



TABLE 1  
VOTING BEFORE AND AFTER FILM EXPOSURE BY CONTROL AND EXPERIMENTAL GROUPS

Vote or selection	Experimental groups		Control group
	Ss told to look for message ( <i>N</i> = 10)	Ss told nothing ( <i>N</i> = 13)	Ss exposed to film ( <i>N</i> = 10)
Message: Vote Wehle			
In favor			
Before film	1	4	5
After film	2	2	3
Against			
Before film	9	9	5
After film	8	11	7
Message: Eat Ybel			
In favor			
Before film	0	2	1
After film	1	4	0
Against			
Before film	10	11	9
After film	9	9	10

tion and those voting in the opposite direction. The chi-square test was applied to averaged data for each group on the potency, evaluative, and activity factors of the semantic differential. Numerical valuations for the seven scale positions were obtained by assigning the digits  $+\frac{1}{2}3$  through  $-3$ .

#### D. SUMMARY

Two experimental groups (one motivated) chose a candidate, selected a product, and made judgments about eight concepts connected with a film to which they were exposed.

Between pretest and posttest, experimental-group Ss were exposed to sub-

TABLE 2  
CHANGES IN AVERAGED FACTOR SCORES AFTER EXPOSURE TO FILM

Group	<i>N</i>	Semantic-differential factor					
		Potency	Activity		Evaluation		
		Party Leader Strong	Press Room Calm	Free Press Passive	Lobbyist Valuable	Legislature Fair	Reporter Bad
Saw film and messages	13	.06	— .62	.14	.32	.25	.24
Saw film and messages and were given warning	10	.17	.13	— .03*	.52	.40	.82
Saw film only	9	— .03	— .14	.16**	.25	— .03	— .48

\* Nine Ss.

\*\* Eight Ss.

liminal messages designed to influence their selections and judgments. A control group saw the film but not the messages.

Changes in voting were largely in the motivated direction; however, the changes found could have occurred by chance.

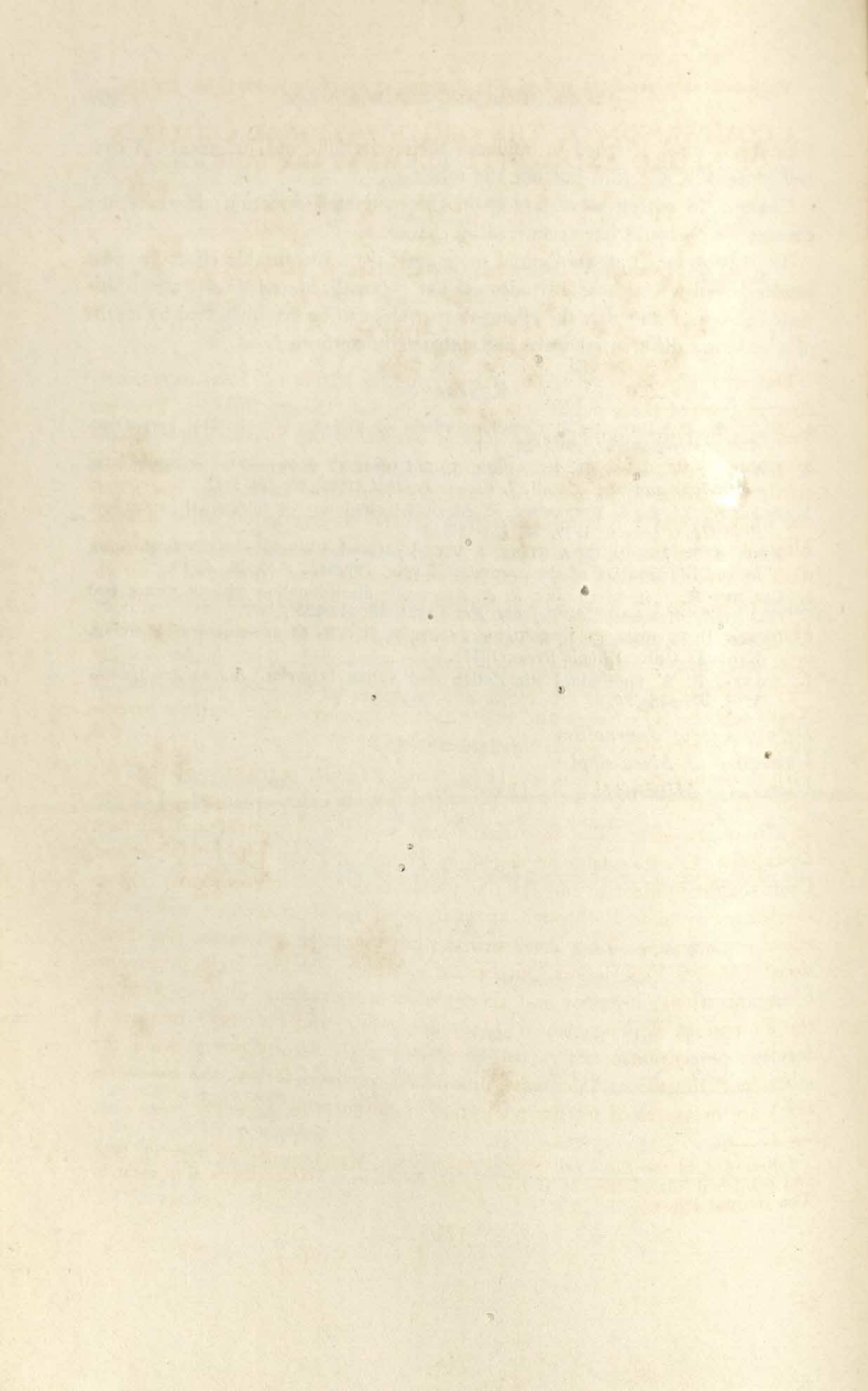
The hypotheses that subliminal messages have a measurable effect upon attitudes (even when those attitudes are not externalized and do not affect ultimate decisions) and that the changed attitudes can be demonstrated by means of a semantic differential were not statistically confirmed.

#### REFERENCES

1. BACH, S., & KLEIN, G. S. Conscious effects of prolonged subliminal persuasion. *Amer. Psychologist*, 1957, **12**, 397.
2. COYNE, J. W., KING, H. E., AUBIN, J., & LANDIS, C. Accuracy of recognition of subliminal auditory stimuli. *J. Exper. Psychol.*, 1943, **33**, 508-513.
3. DEFLEUR, M. L., & PETRANOFF, R. M. A televised test of subliminal persuasion. *Pub. Opin. Quart.*, 1959, **23**, 168-180.
4. KING, H. E., LANDIS, C., & ZUBIN, J. Visual perception where a figure is obscured by the illumination of the ground. *J. Exper. Psychol.*, 1944, **34**, 60-69.
5. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58**, 113-122.
6. OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The Measurement of Meaning*. Urbana: Univ. Illinois Press, 1957.
7. PARKER, E. B. Subliminal stimulation and voting behavior. *Journalism Quart.*, 1960, **37**, 588-590.

*Department of Journalism*  
*University of Mississippi*  
*University, Mississippi*





## A COMPARISON OF THE ORGANIZATIONAL CLIMATES OF NEGRO AND WHITE ELEMENTARY SCHOOLS\*

*College of Education, University of Georgia*

---

HAROLD W. GENTRY AND JAMES B. KENNEY

---

### A. BACKGROUND

During the past 20 years the attention of the nation has been increasingly directed toward those public schools enrolling only Negro children. This has been particularly true of Negro schools located in the South. These schools have been indicted for the low achievement of their graduates, for being primarily vocational in purpose, for being poorly staffed, for having low financial support, and for being inferior in general to segregated white schools. Although considerable research has been directed toward the students and staffs of Negro schools, few empirical studies have been conducted in an attempt to isolate the ways in which these schools are organized and operated. Many questions seem pertinent: for example, Are white schools administered in a more democratic manner than Negro schools? Is there higher morale among faculty members of white schools than among faculty members of Negro schools? Do Negro and white faculty members differ in their perceptions of the schools in which they are employed?

In this study the authors sought to examine faculty perception of the "climate" of both Negro and white elementary schools in an urban school system of a southern state. The instrument used was the Organizational Climate Description Questionnaire developed by Halpin and Croft (4). Halpin and Croft sought to identify and describe the dimensions of organizational climate in elementary schools through an analysis of social interaction within the school organization. They developed the instrument known as the Organizational Climate Description Questionnaire, which identifies eight dimensions of organizational behavior and six types of organizational climate. Four of the dimensions (disengagement, hindrance, esprit, and intimacy) measure a faculty's perception of interaction as related to the faculty group, while the other four dimensions (aloofness, production emphasis, thrust, and consideration) are measures of teacher perception of the principal's behavior.

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on May 12, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



The eight dimensions of organizational behavior, slightly modified by the present authors, were defined by Halpin and Croft (4) as follows:

1. *Disengagement* refers to the teachers' tendency to be "not with it." This dimension describes a group that is "going through the motions," a group that is "not in gear" with respect to the task at hand.

2. *Hindrance* refers to the teachers' feeling that the principal burdens them with routine duties, committee demands, and other requirements which the teachers construe as unnecessary busy work. The teachers perceive the principal as hindering rather than facilitating their work.

3. *Esprit* refers to "morale." The teachers feel that their social needs are being satisfied, and that they are, at the same time, enjoying a sense of accomplishment in their job.

4. *Intimacy* refers to the teachers' enjoyment of friendly social relations with each other. This dimension describes a social-needs satisfaction which is not necessarily associated with task accomplishment.

5. *Aloofness* refers to behavior by the principal which is characterized as formal and impersonal. He "goes by the book" and prefers to be guided by rules and policies rather than to deal with the teachers in an informal, face-to-face situation. His behavior, in brief, is universalistic rather than particularistic; nomothetic rather than idiosyncratic. To maintain this style, he keeps himself—at least, "emotionally"—at a distance from his staff.

6. *Production emphasis* refers to behavior by the principal which is characterized by close supervision of the staff. He is highly directive, and plays the role of a "straw boss." His communication tends to go in only one direction, and he is not sensitive to feedback from the staff.

7. *Thrust* refers to behavior by the principal which is characterized by his evident effort in trying to "move the organization." "Thrust" behavior is marked not by close supervision but by the principal's attempt to motivate the teachers through the example which he personally sets.

8. *Consideration* refers to behavior by the principal which is characterized by an inclination to treat the teachers "humanly," to try to do a little something extra for them in human terms.

Scores on the eight dimensions are patterned to form six types of organizational climate. Ranged along a continuum they are: open, autonomous, controlled, familiar, paternal, and closed. Each climate type is defined as follows (5):

1. *The open climate* describes an energetic, lively organization which is moving toward its goals, and which provides satisfaction for the group members' social needs. Leadership acts emerge easily and appropriately from both the group and the leader. The members are preoccupied disproportionately with neither task achievement nor social-needs satisfaction; satisfaction on both counts seems to be obtained easily

and almost effortlessly. The main characteristic of this climate is the "authenticity" of the behavior that occurs among all members.

2. *The autonomous climate* is described as one in which leadership acts emerge primarily from the group. The leader exerts little control over the group members; high esprit results primarily from social-needs satisfaction. Satisfaction from task-achievement is also present, but to a lesser degree.

3. *The controlled climate* is characterized best as impersonal and highly task-oriented. The group's behavior is directed toward task-accomplishment, while relatively little attention is given to behavior oriented to social-needs satisfaction. Esprit is fairly high, but it reflects achievement at some expense to social-needs satisfaction. This climate lacks openness, or "authenticity" of behavior, because the group is disproportionately preoccupied with task achievement.

4. *The familiar climate* is highly personal, but uncontrolled. The members of this organization satisfy their social needs, but pay relatively little attention to social control in respect to task accomplishment. Accordingly, esprit is not extremely high simply because the group members secure little satisfaction from task achievement. Hence, much of the behavior within this climate can be construed as "inauthentic."

5. *The paternal climate* is characterized best as one in which the principal constrains the emergence of leadership acts from the group and attempts to initiate most of the acts himself. The leadership skills within the group are not used to supplement the principal's own ability to initiate leadership acts. Accordingly, some leadership acts are not even attempted. In short, little satisfaction is obtained in respect to either achievement or social needs; hence, esprit among the members is low.

6. *The closed climate* is characterized by a high degree of apathy on the part of all members of the organization. The organization is not "moving"; esprit is low because the group members secure neither social-needs satisfaction nor the satisfaction that comes from task achievement. The members' behavior can be construed as "inauthentic"; indeed, the organization seems to be stagnant.

## B. PURPOSE AND SETTING

The purpose of this study was to answer two questions: (a) Do the faculties of Negro and white elementary schools in a predominantly segregated-school system perceive the organizational climate of their schools differently? and (b) If the faculties of Negro and white elementary schools perceive the organizational climates of the schools differently, where does the difference lie?

The population of the study included all elementary-school teachers in an urban school district located in a southeastern state. The student enrollment in the school district was rather evenly divided between members of the two



racess. While the school system had complied with a court order to begin integrating Negro and white students, there had been no integration of teaching staffs at the time the study was conducted.

The school system enjoys the prestige of being one of the best in the State. The salary scale is one of the highest in the State and is comparable to those in cities of equal size throughout the Nation. The community in which the school system is located has a history of rather good relations between the races. Negroes serve on the school board and other governmental bodies within the community.

### C. METHOD

The Organizational Climate Description Questionnaire was given to the faculties of all elementary schools in the system during the last quarter of the school year 1964. All questionnaires were completed on the same day. The principal of each school, who had received detailed instruction on the way in which the instrument was to be administered, accepted responsibility for seeing that each teacher in his school completed the questionnaire. The instruments were returned to the central office where they were collected by the investigators. Complete responses were returned from 111 of 114 elementary schools.

To insure anonymity of the participants, no names were required on the questionnaires. After the instruments were completed by faculty members, the instruments were sealed by the principal in a specially prepared envelope in the presence of the faculty and placed in the intrasystem delivery system.

The questionnaires were computer scored with a program provided by the originators of the instrument, and chi squares were computed to test the significance of the difference between the distributions of climate types in each of the two types of schools. A test of significance, using Student's *t*, was employed to identify significant differences between means.

### D. RESULTS

Table 1 shows that the faculties of the Negro and the white elementary schools perceive the organizational climate of the school differently in some instances; in others, similarly. While 40 per cent of the Negro schools were perceived by their faculties as being "closed," only six per cent of the white schools were seen in this manner by their faculties. Approximately 41 per cent of the white faculties saw their school as having an "open" climate, as compared with 13 per cent of the Negro faculties. Thus, in one comparison Negro

TABLE 1  
PERCENTAGE DISTRIBUTION OF ELEMENTARY SCHOOLS BY TYPE OF  
ORGANIZATIONAL CLIMATE AND RACE

Organizational climate	Type of school				Total	
	Negro		White		Number	Per cent
	Number	Per cent	Number	Per cent		
Open	6	13.3	27	40.9	33	29.7
Autonomous	0	—	2	3.0	2	1.8
Controlled	8	17.8	7	10.6	15	13.5
Familiar	0	—	5	7.6	5	4.5
Paternal	13	28.9	21	31.8	34	30.7
Closed	18	40.0	4	6.1	22	19.8
Total	45	100.0	66	100.0	111	100.0

\* Chi square, Negro—white = 28.26,  $p < .01$ .

schools are quantitatively described as tending toward the "closed" end of the continuum, while white schools are more often seen as "open."

On the other hand, 29 per cent of the Negro faculties and 32 per cent of white faculties perceive their school as having a "paternalistic" climate; therefore, when considering all schools irrespective of race, "paternalism" becomes the most frequently selected climate type (30.6 per cent) followed closely by the "open" climate (29.7 per cent) and then by "closed" climate (19.8 per cent).

Table 2 presents a comparison of means of each of the climate dimensions for Negro and white schools by climate type. It should be recalled that the dimension pattern forms six climates: for example, an "open" climate is char-

TABLE 2  
COMPARISON OF DIMENSION MEANS OF NEGRO AND WHITE ELEMENTARY SCHOOLS BY  
TYPE OF ORGANIZATIONAL CLIMATE

Dimension	Organizational climate*					
	Open		<i>t</i>	Controlled		<i>t</i>
	Negro school ( <i>N</i> = 6)	White school ( <i>N</i> = 27)		Negro school ( <i>N</i> = 8)	White school ( <i>N</i> = 7)	
Disengagement	44.3	41.5	2.63**	47.6	42.5	1.70
Hindrance	40.0	42.6	1.09	49.4	50.6	.33
Esprit	49.0	58.9	2.90***	45.2	49.2	.94
Intimacy	44.0	46.2	.82	38.2	38.5	.11
Aloofness	40.3	42.4	1.09	49.2	49.8	.15
Production emphasis	60.0	48.6	4.71***	64.5	61.1	1.94
Thrust	58.8	56.4	1.82	52.4	51.2	.48
Consideration	59.3	62.0	1.10	49.4	52.8	1.08

\* Authoritarian and "familiar" climates not included because no Negro schools fall in these categories.

\*\* Significant at the .05 level.

\*\*\* Significant at the .01 level.



TABLE 2 (continued)

Dimension	Organizational climate*					<i>t</i>
	Paternalistic		Closed			
	Negro school ( <i>N</i> = 13)	White school ( <i>N</i> = 21)	Negro school ( <i>N</i> = 18)	White school ( <i>N</i> = 4)		
Disengagement	58.3	54.9	1.16	59.8	61.2	0.54
Hindrance	47.4	51.1	1.92	51.2	53.8	1.07
Esprit	41.5	42.9	.80	38.4	38.0	.17
Intimacy	39.7	42.5	1.93	44.0	44.0	.03
Aloofness	44.2	41.9	1.07	54.4	48.8	1.10
Production emphasis	56.5	52.3	1.94	59.9	58.2	.52
Thrust	51.8	49.5	.99	41.7	40.2	.57
Consideration	56.5	60.0	1.85	47.9	54.0	2.41**

acterized by high esprit, thrust and consideration and low disengagement, hindrance and production emphasis. It will be noted that significant differences between the two groups occur only at the extremes of the climate continuum. In schools classified as "open" by faculties of both races, Negro faculties perceive themselves as being significantly lower on esprit and significantly higher on disengagement and production emphasis than do faculties of predominantly white schools. On the other hand, in schools with "closed" climates, the white faculties perceive the principal as being significantly more considerate than do Negro faculties.

### E. CONCLUSIONS AND DISCUSSION

The results of the study provide evidence on which answers to the two questions posed may be based. First, Negro and white elementary-school faculties in a predominantly segregated-school system, while in agreement as to the "paternal" climate existing in some schools of both races, generally view the organizational climate of their schools differently. Second, Negro faculties tend to view their schools as falling on the "closed" end of the climate continuum (paternalistic or closed), while white teachers see their school as having a bipolar distribution on the climate continuum (open or paternalistic).

When the analysis is made in terms of dimension scores, the two groups show significant differences in perception only at the extremes of the climate continuum. It is apparent that Negro faculties see the faculty group as having rather low morale and as being highly disengaged from their tasks. At the same time they view the principal as emphasizing production, with modest consideration for the faculty. On the other hand, white faculties tend to have higher morale within the faculty group and to view the principal as hard

working and considerate. Leadership in the Negro faculties is apparently centered in the principal, while in white faculties it arises perhaps both from the faculty group and the principal.

The fact that a relatively high percentage of both white and Negro teachers see their schools as "paternalistic" is of interest. This may be a feature common to the southeastern region. In their original study in which the Organizational Climate Description Questionnaire was developed, Halpin and Croft (4) reported that six out of 10 schools in another state from this region could be classified as "paternalistic." Although the two studies do not offer sufficient data on which to reach a conclusion, they indicate a point of departure for future studies of this nature. In view of the past history of paternalism in society and government within the southeastern region, it seems not unreasonable to expect that many public schools would reflect a "paternalistic" climate resulting from paternalistic leadership.

Why do Negro teachers tend to perceive their schools as "closed," while white teachers tend to perceive a substantial number of their schools as "open"? There are many factors, such as needs satisfaction, values, and cultural background, which affect perception (2). Undoubtedly these factors and many others were operative to a greater or lesser degree in this particular instance. The authors conjecture that one of the more tenable explanations of the findings of this study lies in the role of the principal in the school and community.

Some authorities (1, 3) believe that the principal is the key figure within a school organization. His approach to making decisions, controlling students, developing public relations, and working with the faculty decidedly influence the interactions that occur within the faculty group, the relationship between the school and the central administration, and the relationship between the school and the community. Because of this influence, the principal occupies a sometimes anxious, pressure-ridden, and conflict-prone position. Subject to the wishes, expectations, and pressures of his faculty on one hand and those of the superintendent, school board, and community on the other, the principal (regardless of his race) can truly be characterized as the "man in the middle." Caught in the middle, the principal must find some way to resolve the conflicts and pressures that beset him from both sides. In this respect, the Negro principal probably finds himself in a more untenable position than the principal of a white school because he must mediate between two positions that are more widely separated than those faced by the white principal. In addition, the Negro principal usually occupies a higher-status position



within his school community than does his white counterpart. This circumstance increases the pressure on the Negro principal because he is often placed in a position of representing the wishes of the community, as well as the school faculty, in dealing with the superintendent and board of education.

Faced with the pressures of his job and the conflict of representing two often-divergent points of view, the principal must take some action to resolve his dilemma. What alternatives, then, are open to him? Halpin and Winer (6), in a study involving a military organization suggested that the man in the middle may take several courses of action: (a) he can identify completely with his superiors and the formal organization and disparage the need to be considerate of his subordinates, (b) he can reduce intimacy with his men to minimize guilt feelings about being inconsiderate of them, (c) he can be inconsiderate "on the job" but very considerate of the men off the job, and (d) he can identify completely with his men. If a leader fully accepted any of these positions, he probably would be ineffective due to a failure to fully serve either his superiors or his subordinates. An effective leader would be able to achieve balance in his behavior, so that his superior and his subordinates will not view him as being "the man" of the other group.

Following this line of thinking, the data presented in this paper seem to indicate that white principals have achieved a degree of balance in their operation. Morale in their schools is generally high, the teachers view the principal as considerate, and there seems to be a reasonable balance between production emphasis and thrust. In other words, the organization apparently meets the twin needs of the faculty group: e.g., social satisfaction and task accomplishment. Conversely, the Negro principal appears to have become more oriented to his superiors with a corresponding lessening of concern for his subordinates. While working hard himself and emphasizing production, there apparently is low task accomplishment due to low morale and high disengagement on the part of the faculty.

#### F. SUMMARY

In this study, the authors sought (a) to determine if faculties of Negro and white elementary schools in a predominantly segregated-school system perceive the organizational climate of their schools differently and, if they do, (b) to determine where the difference lies.

The Organizational Climate Description Questionnaire developed by Halpin and Croft was used as the data-gathering instrument. The sample consisted of 111 elementary schools in an urban school district in a South-eastern State. Forty-five of the sample schools were Negro, 66 were white.

Negro and white faculties differ significantly in their perception of the organizational climate of their schools. A large majority of Negro faculties perceive their schools as either "closed" or "paternalistic" (as defined by the model presented), while white faculties see their schools as "open" or "paternalistic."

## REFERENCES

1. BRIDGES, E. M. Teacher participation in decision making. *Administrator's Notebook*, 1964, 12, 1-4.
2. COSTELLO, G. W., & ZALKIN, S. S. *Psychology in Administration*. Englewood Cliffs: Prentice Hall, 1964.
3. GRIFFITHS, D. E. *Human Relations in School Administration*. New York: Appleton, 1956.
4. HALPIN, A. W., & CROFT, D. B. The Organizational Climate of Schools. Contract Number SAE543 (8639), United States Office of Education, Washington, D. C., July, 1962.
5. ———. The organizational climate of schools. *Administrator's Notebook*, 1963, 11, 1-4.
6. HALPIN, A. W., & WINER, B. J. A factorial study of the leader behavior descriptions. In R. M. Stogdill and A. E. Coons (Eds.), *Leader Behavior: Its Description and Measurement*. Bureau of Business Research Monograph (Ohio State University), 1957, 88. Cited by B. M. Bass, *Leadership, Psychology and Organizational Behavior*. New York: Harper, 1960. P. 319.

College of Education  
The University of Georgia  
Athens, Georgia





## THE RELATIONSHIP BETWEEN EXPRESSED SELF-ESTEEM AND ASSUMED SIMILARITY\*<sup>1</sup>

*Department of Psychology, Marymount Manhattan College*

MAUREEN P. FITZGERALD<sup>2</sup>

### A. INTRODUCTION

In social interaction, each individual puts forth some claims as to what sort of person he is and what sort of behavior he expects from others. The anticipated behavior is affected by (a) the image the individual has of himself, (b) the image he has of others, and (c) the image he believes others have of him. Heider (8); Kogan and Tagiuri (9); Kogan, Tagiuri, and Portis (10); and Tagiuri, Blake, and Bruner (14) stress the reciprocity of these relations. Festinger (3, 4) would cast these relations in terms of consonance amongst relevant cognitions. \*

The attitudes reflected in assumed similarity (i.e., the degree to which one person perceives another as similar to himself) are attitudes of which the individual is frequently unaware. High assumed similarity supposedly indicates acceptance of others; while a low assumed-similarity score is considered more indicative of a rejecting, psychologically distant attitude (5). Investigators have sought to relate assumed similarity or social distance to adjustment and effective interpersonal relations.

Previous studies employing normal populations have considered high self-esteem or self-acceptance to be a trait common in all their Ss (1, 6, 8, 11, 12, 13). It would seem that the degree of self-esteem would affect the similarity assumed to others. The present study consisted of an attempt to examine the foregoing assumption more closely and to determine the degree to which self-esteem and assumed similarity are related. Self-esteem is herein defined as the rating a person gives himself on a set of items descriptive of per-

\* Received in the Editorial Office, Provincetown, Massachusetts, on May 17, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> A summary of a portion of these findings was presented at the meetings of the Eastern Psychological Association, Philadelphia, 1964. The work was performed as part of a Doctoral dissertation under the direction of Professor Joseph F. Kubis, Fordham University.

<sup>2</sup> Many thanks are due to J. Keegan, S.J., A. Schneiders, R. Zlody, Mary Reuder, and especially to W. E. Collins for constructive suggestions stemming from readings of this manuscript in various stages of preparation.



sonal attributes (6). Three main hypotheses are tested. The first compares the similarity between the image the individual has of himself and the image he has of others and its relationship to self-esteem. The second and third hypotheses attempt to extend the concept of assumed similarity and consonant relations to the similarity between the image the *S* has of himself and the image he believes others have of him.

First, it was predicted that the magnitude of the discrepancy or dissonance between the image of the self and the image the *S* has of another—an average person, one who is neither a close friend nor one disliked—would be greater for those expressing low self-esteem than for those expressing high self-esteem. This hypothesis stemmed from the consideration that what a person expects from his world is in part determined by his own personality. The person with high self-esteem has many cognitions favorable to himself and relatively few that are unfavorable. In other words, he "lives" in a friendlier environment; therefore, he perceives positive traits in himself and others. The person with low self-esteem evidences fewer positive traits and, in experiencing the world as more thwarting and threatening, perceives others as psychologically more distant; hence less similar.

Second, it was predicted that the magnitude of the dissonance or discrepancy between a self-rating and an estimate of a rating of the self by one liked best would be greater for those expressing low self-esteem than for those expressing high self-esteem. If a high self-esteem *S* was asked to estimate how one he liked best would rate him, it was expected that he would assume that his own self-accepting view would be reciprocated. A person of low self-esteem, feeling himself less acceptable to others, could either anticipate that the one liked best does not return this liking or that a genuine reciprocated liking could be attributed to the fact that the *S* was incorrectly perceived as having more positive traits than he believes he actually has.

Third, it was predicted that a person expressing high self-esteem would expect a rating from one liked least to be negative but not so negative as the expectancy of a low self-esteem *S* in the same situation. Heider (8); Kogan and Tagiuri (9); Kogan, Tagiuri, and Fortis (10); and Tagiuri, Blake, and Bruner (14) offer evidence that persons expect to be disliked by those they in turn dislike—hence those expressing both high self-esteem and low self-esteem would expect a negative rating from a person liked least. Yet even in this unfavorable situation the person of high self-esteem sees himself as being more adequate in terms of cultural acceptance than does the person of low self-esteem. It should be noted that according to dissonance theory in this in-

stance neither dissonance nor consonance would exist because a rating of the self by one liked least is not a *relevant* cognition (2, 5).

Special provisions were made to analyze the data of a middle group: i.e., those expressing neither high self-esteem nor low self-esteem, but rather a median amount of self-esteem. Too often in experimentation untested assumptions are made concerning a middle group without any attempt at verification. The question as herein conceived was, "Is the middle group really a middle group in the amount of social distance maintained or assumed to exist in interpersonal relation?"

To summarize, three operationally defined predictions are made: Female Ss expressing high self-esteem, in comparison with those expressing low self-esteem, (a) show significantly greater similarity to an "average girl," (b) estimate significantly greater similarity between a self-rating and an estimate of the rating of the S by the girl liked best, and (c) estimate greater similarity between a self-rating and an estimate of the rating of the S by the girl liked least.

## B. METHOD

### 1. Subjects

The 300 Ss were unmarried white female day students at a small denominational college for women in New York. There were 75 freshmen, 75 sophomores, 75 juniors, and 75 seniors, most of them of middle-class socioeconomic families in the metropolitan area. As it is the policy of the college to have a series of events to integrate each student actively into the college life, by December of the freshman year each student knows every other member of her class by name and by sight and has had some interaction with each of them.

### 2. Procedure

For an index of expressed self-esteem, each S completed a self-rating scale on which the S was requested to "describe yourself as you ordinarily think about yourself." The scale consisted of 20 bipolar pairs of adjectives scored on a good-bad continuum. The scale and scoring procedure were those used by Fiedler *et al.* (6) and Fitzgerald (7). The highest possible score was 120. The upper, middle, and lowest 20 per cent of each college class were selected on the basis of total score to constitute the high-self-esteem, the median-self-esteem and the low-self-esteem groups respectively; thus 15 Ss constituted each of the self-esteem groups.



### 3. Scores

Three assumed-similarity scores were computed.

*a. Assumed similarity to average.* Following the self-rating scale, the same form and rating procedure were used to obtain the *S*'s description of the "average girl in your class, one who is neither a close friend of yours, nor one disliked by you." Essentially Cronbach and Glaser's *D* statistic [see (6)], which measures the profile similarity between two ratings, was used for this index and for the two following. To compute the assumed-similarity-to-the-average-girl score for each *S*, the self-rating scale and the rating of the average or typical girl in the class were compared. The difference between ratings for each item on the two forms was computed and squared. The square root of the summed squared differences multiplied by 10 became the *D* score.

*b. Similarity to estimate of liked best.* Using the same rating form, the *S*'s estimate of the image others had of her was obtained. Each *S* was asked "to think of the girl in your class whom you like best. Now describe yourself as this girl sees you, what she thinks of you." A *D* score was computed between the self-rating and the *S*'s estimate of her rating by the girl in her class liked best.

*c. Similarity to estimate of liked least.* The *S*s were also asked to "think of the girl in the class whom you like least. Describe yourself as this girl sees you, what she thinks of you." The *D* statistic was computed between the self-rating and the *S*'s estimate of her rating by the girl in the class liked least.

The experimental design enabled the testing of two subsidiary hypotheses. First, it was not expected that freshmen, sophomores, juniors, and seniors would differ in the amount of similarity assumed to others; yet each of the four classes was tested separately, and the results were analyzed separately to detect any differences that might arise. Second, the question was posed: Does the median-self-esteem group fall midway between the high-self-esteem group and the low-self-esteem group in the amount of similarity assumed to others? Both of the subsidiary hypotheses were investigated for each of the three main hypotheses.

## C. RESULTS

### 1. Assumed Similarity to Average

Figure 1 shows the mean discrepancy scores for assumed similarity to the average girl in the class for the low-self-esteem, median-self-esteem, and high-self-esteem groups. The larger the *D* score, the greater is the discrepancy or dissonance between the *S*'s self-rating and her rating of the average girl in the

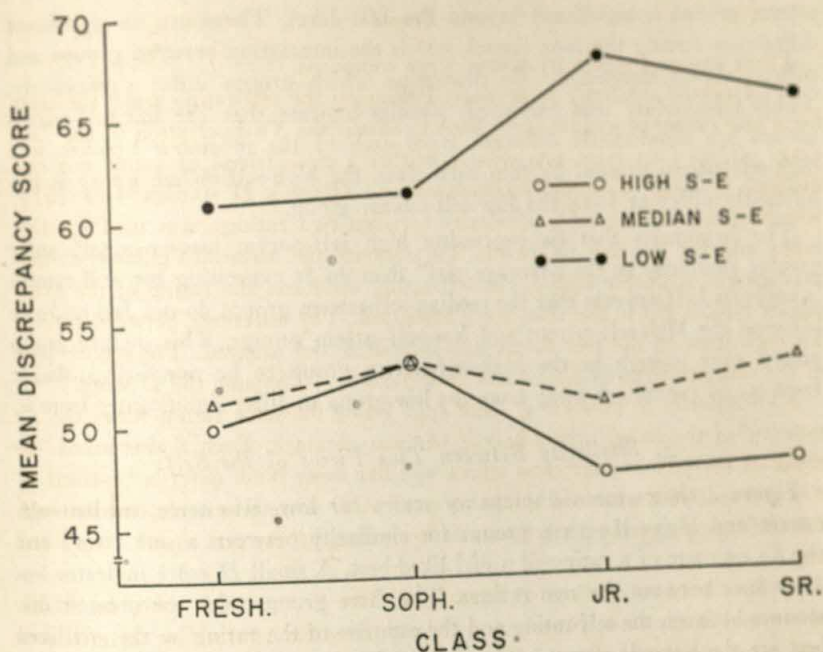


FIGURE 1  
MEAN DISCREPANCY SCORES FOR ASSUMED SIMILARITY TO THE AVERAGE GIRL  
FOR THREE SELF-ESTEEM GROUPS

class. The smaller the score the greater is the similarity assumed to this average girl. The four low-self-esteem groups assumed the least amount of similarity, and three of the high-self-esteem groups assumed the greatest amount of similarity to this average girl.

Table 1 shows that the  $F$  ratio of 11.1 for differences among the self-

TABLE 1  
ANALYSIS OF VARIANCE FOR ASSUMED SIMILARITY TO THE AVERAGE  
GIRL FOR THREE SELF-ESTEEM GROUPS

Source	df	Sum of squares	Mean square	F
Between groups	11	8,205.4	745.95	2.29*
Error term	168	54,680.2	325.48	11.1***
Between self-esteem groups	2	7,225.93	3,612.96	.116
Between college classes	3	114.08	38.02	.44
Interaction	6	865.39	144.23	

\* Significant at the .05 level.

\*\*\* Significant at the .001 level.



esteem groups is significant beyond the .001 level. There are no significant differences among the four classes, nor is the interaction between groups and college classes significant. To determine which groups differ significantly, Tukey's procedure was employed. Results indicate that the low-self-esteem groups are significantly different from each of the median-self-esteem and high-self-esteem groups. Within each class, the high-self-esteem group is significantly different from the low-self-esteem group.

The hypothesis that *Ss* expressing high self-esteem assume significantly greater similarity to the "average girl" than do *Ss* expressing low-self-esteem is supported. It appears that the median-self-esteem groups do not fall midway between the high-self-esteem and low-self-esteem groups. This middle group placed near enough to the high-self-esteem group to be nondistinguishable from it and far enough away from the low group to differ significantly from it.

## 2. *Similarity Between Two Views of the Self*

Figure 2 shows mean discrepancy scores for low-self-esteem, median-self-esteem and high-self-esteem groups for similarity between a self rating and the *S's* estimate of a rating of a girl liked best. A small *D* score indicates less dissonance between the two ratings. The three groups with the greatest dissonance between the self-rating and the estimate of the rating by the girl liked best are the low-self-esteem groups, while the three groups with the least dissonance between ratings are the high-self-esteem groups. Table 2 shows the *F* ratio of 15.35 for differences among the self-esteem groups to be significant beyond the .001 level. There are no significant differences among the four classes. The high-self-esteem freshmen, sophomores, and seniors are significantly different from all four low-self-esteem groups; and all four high-self-esteem groups are significantly different from the low-self-esteem freshmen, sophomores, and seniors. The median-self-esteem groups differentiate themselves from *some* of the high-self-esteem groups but no median-self-esteem group differentiates itself from the high-self-esteem groups of its own class. There is some tendency for the median-self-esteem groups to fall nearer to the high-self-esteem groups and further away from the low-self-esteem groups. Three of the four classes support the hypothesis that *Ss* expressing high-self-esteem evidence less dissonance between a self-rating and an estimate of their rating by a girl liked best than do those expressing low-self-esteem.

To test further the similarity between the self-rating and the *S's* estimate of another's rating of her, data pertaining to the girl liked least were analyzed. A plot of the *D* scores between these two ratings is given in Figure 3.

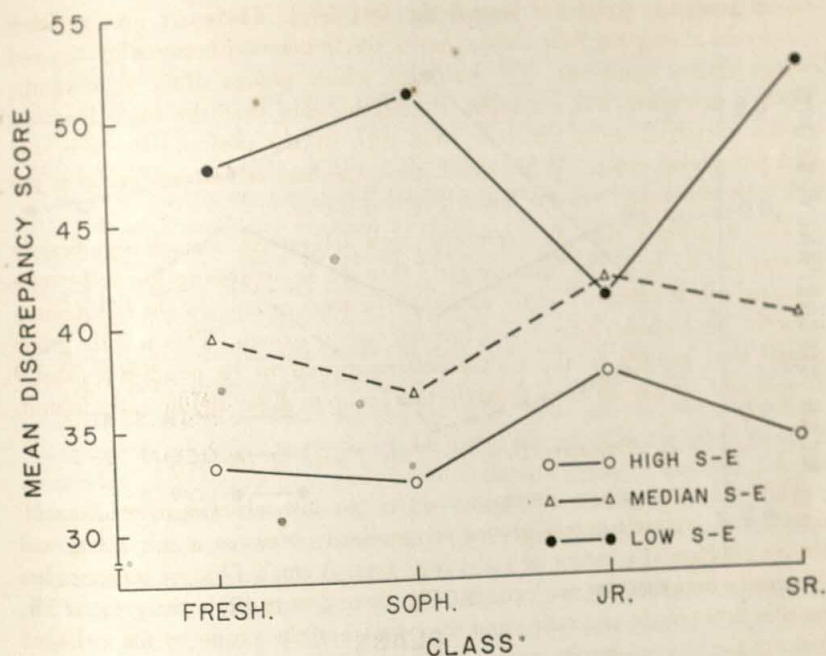


FIGURE 2

MEAN\* DISCREPANCY SCORES FOR SIMILARITY BETWEEN SELF-RATING AND ESTIMATE OF RATING OF THE SUBJECT BY A GIRL LIKED BEST FOR THREE SELF-ESTEEM GROUPS

Three of the four low-self-esteem groups have the largest  $D$  scores, but only two of the high-self-esteem groups have the lowest  $D$  scores. The two lowest  $D$  scores are obtained by the median-self-esteem groups. The median-self-esteem and low-self-esteem seniors yield almost the same amount of dis-

TABLE 2  
ANALYSIS OF VARIANCE OF SIMILARITY BETWEEN SELF-RATING AND ESTIMATE OF RATING OF SUBJECT BY GIRL LIKED BEST FOR THREE SELF-ESTEEM GROUPS

Source	df	Sum of squares	Mean square	F
Between groups	11	7,525.1	684.1	3.54*
Error terms	168	32,420.3	192.98	
Between self-esteem groups		5,924.76	2,962.38	15.35***
Between college classes		144.64	48.21	.249
Interaction		1,455.7	242.61	1.257

\* Significant at the .01 level.  
\*\*\* Significant at the .001 level.



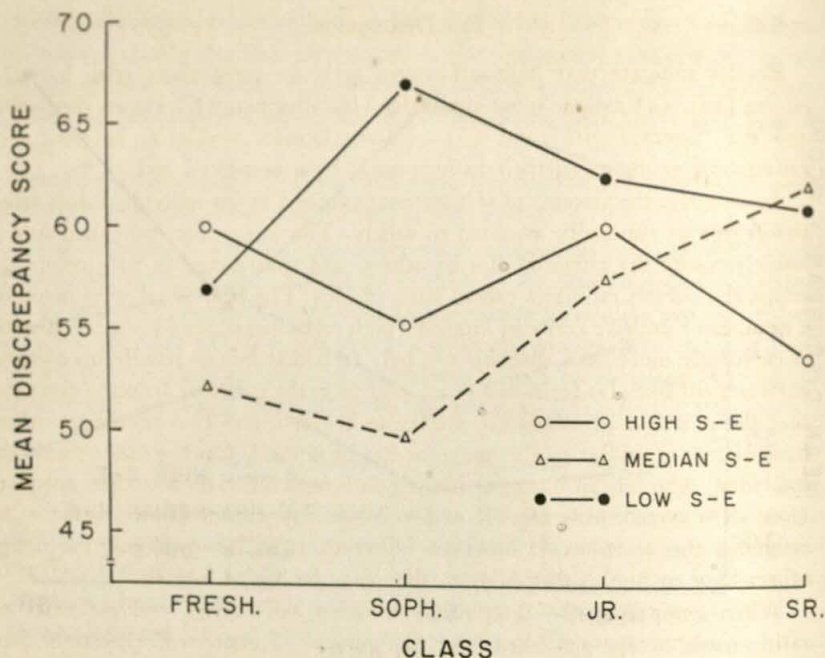


FIGURE 3

MEAN DISCREPANCY SCORES FOR SIMILARITY BETWEEN SELF-RATING AND ESTIMATE OF RATING OF THE SUBJECT BY A GIRL LIKED LEAST FOR THREE SELF-ESTEEM GROUPS

nance between their two ratings (61.9 and 61.0). The junior high-self-esteem and median-self-esteem groups have *D* scores that are similar—60.0 and 57.5 respectively. The analysis of variance [see Table 3] yields insignificant *F* ratios; therefore the hypothesis that *Ss* expressing high-self-esteem evidence greater similarity between a self-rating and an estimate of their rating by a girl liked least than do *Ss* expressing low-self-esteem is not supported.

TABLE 3

ANALYSIS OF VARIANCE OF SIMILARITY BETWEEN SELF-RATING AND SUBJECT'S ESTIMATE OF RATING OF SUBJECT BY GIRL LIKED LEAST FOR THREE SELF-ESTEEM GROUPS

Source	<i>df</i>	Sum of squares	Mean square	<i>F</i>
Between groups	11	3,910.5	355.5	.915
Error term	168	65,268.9	388.51	
Between self-esteem groups	2	1,346.08	673.04	1.73
Between college classes	3	331.60	110.53	.44
Interaction	6	2,232.82	372.14	.96

## D. DISCUSSION

Results indicate that high-self-esteem girls, in comparison with low-self-esteem girls, (a) assume more similarity (less dissonance) between themselves and the "average girl" and (b) evidence greater similarity between self-ratings and estimates of their ratings made by a best-liked girl.

Apparently, the amount of self-esteem expressed by an individual does affect the degree of similarity assumed to others. The image the individual has of himself colors the image he has of others, and that image in turn colors the image the person estimates others have of him. The individual who perceives a number of positive traits in himself (high self-esteem) and perceives himself as culturally more than adequate not only feels that he can handle his cultural environment but also feels, in a sense, that he *is* the culture; hence he estimates that the reactions of others are similar to his own, and that the image others have of him is similar to the image he has of himself. On the other hand, the individual who has an image of himself as being culturally less than adequate (low self-esteem) feels himself as psychologically distant from others; so he estimates the reactions of others as different from his own and the image others have of him as dissimilar to the image he has of himself.

When comparing the discrepancy between self-ratings and estimates of rating made by the girl liked least, the amount of expressed self-esteem does not differentiate the groups. This fact corroborates dissonance theory, which implies that there is neither consonance nor dissonance between a cognition represented by a self-rating and a cognition represented by an estimate of the self by one liked least. For the *S*, the view one liked least has of her is neither sufficiently relevant nor meaningful to cause dissonance (3, 4). It should also be noted that dislikes are less easily revealed in a testing situation. Perhaps, even to the self, attitudes toward one disliked are difficult to clarify. Heider (8); Tagiuri, Blake, and Portis (14); Kogan and Tagiuri (9); and Kogan, Tagiuri, and Portis (10) generally have found too few cases of expressed dislike to test adequately hypotheses pertaining to those disliked. In the very active attempts to reduce prejudices, negative attitudes toward others become associated with that which is unwholesome; hence analysis of or speculation about these negative attitudes toward others by the individual holding them may be hastily pushed from the mind. It is probably easier for an individual in a testing situation (and to himself) to say "There is no one I dislike" than to tap attitudes that may arouse guilt feelings.

The two extreme-self-esteem groups in this study are differentiated statistically, but the median-self-esteem group is not clearly a middle group. In



social distance assumed to the average girl in the class, the median-self-esteem group is nondistinguishable from the high-self-esteem group. With reference to the social distance between self-rating and the S's estimate of her own rating by a girl liked best, none of the median-self-esteem groups is significantly different from the two extreme-self-esteem groups. Beyond a certain point, higher levels of expressed self-esteem appear unrelated to increases or decreases in the amount of social distance maintained or assumed to exist. However, Ss at lower-self-esteem levels feel less similar to and less accepted by others.

### E. SUMMARY

An attempt was made to specify expressed self-esteem as an antecedent variable affecting the magnitude of assumed similarity. Results indicate that Ss high in self-esteem assume more similarity between themselves and the average girl and evidence greater similarity between self-ratings and estimates of their ratings than do Ss low in self-esteem. The amount of expressed self-esteem does not affect the discrepancy between self-ratings and estimates of ratings made by the girl liked least.

### REFERENCES

1. ALPERT, E. Two components of assumed similarity. *J. Abn. & Soc. Psychol.*, 1958, **56**, 135-138.
2. BREHM, J., & COHEN, A. Exploration in Cognitive Dissonance. New York: Wiley, 1962.
3. FESTINGER, L. Theory of Cognitive Dissonance. Stanford: Stanford Univ. Press, 1962.
4. FESTINGER, L., & BRAMEL, D. The reactions of humans to cognitive dissonance. In A. F. Bachrach (Ed.), *Experimental Foundations of Clinical Psychology*. New York: Basic Bks., 1962.
5. FIEDLER, F. Interpersonal perception and group effectiveness. In R. Tagiuri & L. Petrullo (Eds.), *Person Perception and Interpersonal Behavior*. Stanford: Stanford Univ. 1958. Pp. 243-257.
6. FIEDLER, F., HUTCHINS, E., & DODGE, J. Quasi-therapeutic relations in college and military groups. *Psychol. Monog.*, 1959, **73**, No. 3 (Whole No. 473).
7. FITZGERALD, M. Self-disclosure and expressed self-esteem, social distance and areas of the self revealed. *J. of Psychol.*, 1963, **56**, 405-412.
8. HEIDER, F. The Psychology of Interpersonal Relations. New York: Wiley, 1958.
9. KOGAN, N., & TAGIURI, R. Interpersonal preference and cognitive organization. *J. Abn. & Soc. Psychol.*, 1958, **56**, 113-116.
10. KOGAN, N., TAGIURI, R., & PORTIS, B. Perception of reciprocity and the grouping principle. *J. Soc. Psychol.*, 1959, **48**, 27-32.
11. MARTIRE, J., & HORNBERGER, R. Self-acceptance of sex and between sexes in a normal population. *J. Clin. Psychol.*, 1957, **13**, 288-291.
12. MCKENNA, H., HOFSTAETTER, P., & O'CONNOR, J. Conception of the ideal and the friend. *J. Personal*, 1956, **24**, 262-271.

13. SMITH, G. Six measures of self-concept discrepancy and instability of their interrelations, reliabilities, and relation to personality measured. *J. Consult. Psychol.*, 1958, 22, 101-112.
14. TAGIURI, R., BLAKE, E., & BRUNER, J. Some determinants of the perception of positive and negative feelings in others. *J. Abn. & Soc. Psychol.*, 1953, 48, 583-593.

*Department of Psychology*  
*Marymount Manhattan College*  
*221 East Seventy-First Street*  
*New York 21, New York \**





## ADAPTIVE TESTING IN AN OLDER POPULATION\*<sup>1</sup>

*Colorado State Hospital and The Pennsylvania State University*

DENNIS I. GREENWOOD AND CHARLES TAYLOR

### A. INTRODUCTION

The intelligence testing of older people presents many special problems, such as obtaining an appropriate test instrument, eliciting cooperation, and providing for conditions that help to insure that subjects are working under appropriately strong motivation (9). Commenting on the similar difficulty of involving young children, Jones (6) suggested that "beyond the age of 60, it is reasonable to suppose that motivational losses may account in part for the decline in mental ability scores." Pressey (14), in urging tests suited to older people, indicated that "there should be motivation appropriate to the old, and careful avoiding of threat to status or self-expectation."

The unwillingness of older people to participate in mental testing has been ascribed to their low motivational levels (6, 10). On the contrary, Welford (18) concluded that unwillingness to be tested seems to stem from the fear of doing badly and appearing foolish and may be a case of high emotional involvement.

Kay (8) has suggested that we do not know whether changes in learning ability affect motivation or whether changes in motivation themselves affect learning ability. Longitudinal studies that have shown no essential decline of intelligence with age (2, 13) have been based upon subjects of high ability, with relatively little to fear from the testing (6). Could the seeming deterioration shown in cross-sectional studies (7, 12, 16) be lessened by adapting an intelligence scale to permit scoring on the same questions but providing for a quieting of anxiety? A modification of testing procedure described as "adaptive" testing is especially designed to maintain good rapport and seems suited to such an effort. To insure an initial success, the examiner begins the test with an item below the subject's anticipated mental level: i.e., with an item that does not require considerable concentration, rapid response, or prolonged and involved verbal directions. Hard and easy items are spaced to avoid suc-

\* Received in the Editorial Office, Provincetown, Massachusetts, on May 18, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This paper is based on a master's thesis submitted by the first author to Brigham Young University. The second author directed the research and prepared this report.



cessive failures as items become more difficult. Occasionally, to build self-confidence, the subject is given an easy nonscored item. Hutt (5), using the Stanford-Binet, found that this method appreciably raised the performance level of maladjusted children but not of normal children.

The standard procedure of giving the items of each subtest of the Wechsler Adult Intelligence Scale in consecutive order from "easy" to "hard" offers the subject a task of constantly increasing difficulty and requires him to be able to tolerate the increasing frustration inherent in such a situation, even though new subtests start the person once again at a point at which questions are easier.

It is suggested that the older person, in the test situation at least, resembles the less-adjusted children of Hutt's study. The older person has less confidence in his own judgments and reacts better with consistent feedback (15). If the emotional reactions reported from testing older subjects might be allayed in some way, scores might reflect more fairly their capabilities. Although no literature exists in which the Wechsler scales have been used in adaptive testing, in the main they may be cast in that mold.

To convert the WAIS to adaptive testing, the following principles can be used:

1. Each subtest of the WAIS is begun with an item below the anticipated mental level of the subject in order to insure an initial success with the item.
2. The test is started with general information, presumably requiring little concentration or a rapid response.
3. After the subject has failed an item from the test, he is given an easier item from the same subtest. A number of nonscored items must be prepared to use when the pool of easy items has been exhausted.
4. The subtests that do not lend themselves to the adaptive procedure are given by standard procedure: for example, object assembly and digit symbols. Because retests with the object-assembly test lose much of their discriminatory value, this subscale must be omitted from the first testing; and the scaled scores of the other performance subtests prorated to determine the performance score. On a retest, object assembly can be given by the standard procedure.

## B. PROCEDURE

As subjects, 30 volunteers (ages 65 to 75) were solicited at church meetings. Two groups were formed. Using the results of a first testing with the Wechsler Adult Intelligence Scale, the subjects were paired within five years of chronological age and within five *IQ* points. A table of random numbers

was employed to assign 15 subjects to each group. Table 1 shows the nature of the two groups.

TABLE 1  
COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS

Group	Mean age	Original mean IQ	Mean years of education
Experimental	69.1	109.9	11.1
Control	68.5	110.4	11.2

After an average time lapse of three months, the experimental group was retested with the adaptive method; the control group, with the consecutive method.

### C. RESULTS

The Walsh test, a nonparametric statistical test, was used to determine the significance of the difference between the means. This technique is based on difference scores for even pairs of matched subjects (with the difference between those differences arranged in order of size) and with the sign of each difference taken into consideration. Table 2 shows the difference between experimental-group and control-group differences. The statistic shows that the difference between the two group differences is significant beyond the .01 level.

To test whether or not the adaptive testing procedure would produce changes in the same direction in a younger population, the same experimental procedure was used on two groups of college students, 10 each in the experimental and control groups. The original mean IQ of the student groups was 7.4 points higher than the mean of the adult groups. The difference between the original and experimental conditions showed a rise of 1.8 points, and the mean rise upon retesting in the control group was 1.9 points. The Walsh test of course showed no significant increase from the experimental retest over that of test practice in the control group. Presumably, as in Hutt's study, the adaptive method serves to allay test disquietude only in an anxious population.

TABLE 2  
DIFFERENCES BETWEEN EXPERIMENTAL AND CONTROL GROUP TEST-RETEST DIFFERENCES\*

Group	Mean original IQ	Mean retest IQ	Test-retest difference
Experimental	109.9	117.1	7.1
Control	110.4	114.1	3.7

\* Mean intergroup difference = 3.5.



Several of the younger groups perceived that hard missed questions were followed by easy items and commented with some irritation on the procedure.

#### D. DISCUSSION

The tendency for the modified method of testing to increase the scores of those subjects in the experimental group to a greater extent than those of the control group must be attributed to something more than test practice. Maximum retest benefits should come from a *de facto* repetition of the test. The attribution of test stress to an older group seems to be warranted. Due to the selection procedure used, the subjects for this study were probably not highly concerned with the threat of failure. A less-confident group might be expected to show a greater tendency for increased scores with the experimental changes.

The subjects used, however, did show considerable reticence in approaching the test. Typical remarks were "Why don't you test younger people? They would do much better" and "You don't want me. I'm not smart enough." In Hutt's study only the "very poorly adjusted" children obtained reliably higher *IQ* ratings, suggesting that older adults may be, at least in the test situation, sufficiently sensitive to anxiety to obscure their best performance (3).

Because the adaptive method requires the examiner to be constantly alert to the subject's success with successive items so that on-the-spot scoring may determine the nature of the next item, rapport with the subject may be enhanced. *E* was aware of this possibility and attempted to achieve maximum rapport with each subject no matter what the method. The subjects, however, seemed to be more at ease during the testing with the adaptive method and made comments about having enjoyed the experience.

Results obtained must be accepted with caution. Comparability with other test data is difficult, as is the case when time elements are manipulated (11). There is no way of knowing how much the validity of the test is affected by this approach. Tests that overly emphasize "fairness" have been shown to lessen their ability to predict the level of adequacy of meeting everyday-life situations (1).

The modest gains made under the adaptive method would certainly not cause one to make significant changes in life plans for older people from scores made by this method. If, however, a combination of factors [such as speed (11), losses in sensory acuity (17), and unfamiliarity of material (4)] lessen the performance of the aged, impressions about the intellectual functioning

of older persons, when the tests are not representative of the daily activities of older persons, may be misleading.

### E. SUMMARY

The intellectual testing of older people presents special problems in the nature of the instrument, developing rapport, and providing for conditions of adequate motivation. Although older people have a desire to do well on tests, there are signs of considerable test anxiety.

The hypothesis was offered that a modified or "adaptive" order of presenting the items of the WAIS yields reliably higher scores on retest than do conventional retest procedures. Thirty subjects matched on age, years of education, and *IQ* were retested after an average of three months: half by conventional methods, half by the adaptive method. The latter technique involved beginning each test with items below the subject's anticipated mental level, then alternating hard and easy items. Easy test items were supplemented by a pool of nonscored items similar to easy items on the WAIS. The increase in scores from using the adaptive method as compared to that from the consecutive method was significant at the .01 level of confidence.

It was suggested that the score obtained in an anxiety-reducing circumstance may be more indicative of performance in most life situations than is the score obtained in conventional test administration.

### REFERENCES

1. ANASTASI, A. *Differential Psychology*. New York: MacMillan, 1958.
2. BAYLEY, N., & ODEN, M. E. The maintenance of intellectual ability in gifted adults. *J. Geront.*, 1955, **10**, 91-107.
3. BYRD, E. Measured anxiety in old age. *Psychol. Reports*, 1959, **5**, 439-440.
4. DEMING, J. A., & PRESSEY, S. L. Tests "indigenous" to adult and older years. *J. Counsel. Psychol.*, 1957, **4**, 144-148.
5. HUTT, M. L. A clinical study of "consecutive" and "adaptive" testing with the revised Stanford-Binet. *J. Consult. Psychol.*, 1947, **11**, 93-103.
6. JONES, H. E. Intelligence and problem solving. In J. E. Birren (Ed.), *Handbook of Aging and the Individual*. Chicago: Univ. Chicago Press, 1959.
7. JONES, H. E., & CONRAD, H. S. The growth and decline of intelligence: The study of a homogeneous group between the ages of ten and sixty. *Genet. Psychol. Monog.*, 1933, **13**, 223-298.
8. KAY, H. Some experiments on adult learning. In *Old Age in the Modern World*. London: E. & S. Livingstone, 1955.
9. KUHNEN, R. G. Problems in the motivation of subjects in research on aging. *Research on Aging*, Social Science Research Council, New York, 1950.
10. ———. Aging and life-adjustment. In J. E. Birren (Ed.), *Handbook of Aging and the Individual*. Chicago: Univ. Chicago Press, 1959.
11. LORGE, I. The influence of the test upon the nature of mental decline as a function of age. *J. Educ. Psychol.*, 1936, **27**, 100-110.



12. MILES, C. C., & MILES, W. R. The correlation of intelligence scores and chronological age from early to late maturity. *Amer. J. Psychol.*, 1932, **44**, 44-78.
13. OWENS, W. A. Age and mental abilities, a longitudinal study. *Genet. Psychol. Monog.*, 1953, **48**, 3-54.
14. PRESSEY, S. L. Major problems—and the major problem—motivation, learning, and education in the later years. In J. E. Anderson (Ed.), *Psychological Aspects of Aging*. Washington, D. C.: American Psychological Association, 1956.
15. WALLACH, M. A., & KOGAN, N. Aspects of judgment and decision making: Interrelationships and changes with age. *Behavioral Sci.*, 1961, **6**, 23-36.
16. WECHSLER, D. The measurement and evaluation of intelligence of older persons. In *Old Age in the Modern World*. London: E. & S. Livingstone, 1955.
17. WEISS, A. D. Sensory functions. In J. E. Birren (Ed.), *Handbook of Aging and the Individual*. Chicago: Univ. Chicago Press, 1959.
18. WELFORD, A. T. Skill and Age: An Experimental Approach. London: Oxford Univ. Press, 1951.

*Continuing Education*

*Pennsylvania State University*

*302 Engineering D*

*University Park, Pennsylvania 16802*

## CONTRAST EFFECTS IN EXTINCTION RESPONDING AS A FUNCTION OF RESPONSE EFFORT DURING CONDITIONING\*<sup>1</sup>

*Department of Psychology, Illinois Wesleyan University*

---

THOMAS STACHNIK

---

### A. INTRODUCTION

Contrast is a well-explored phenomenon in sensory psychology. For example, if a light-gray and a dark-gray square are juxtaposed, the brightness difference between them will seem to be augmented immediately: a phenomenon called brightness contrast. Similarly, each of two colored squares placed side by side will be modified in the direction of the complement of the other; and this is called color contrast. Recently, types of contrast effects have been noted in behavioral studies that in deference to sensory psychology have been called "behavioral contrast" (4). Contrast effects have been generated under a number of experimental procedures. The term behavioral contrast was coined in connection with a discrimination experiment where in a free-operant situation two stimuli (colored lights) are alternated at various intervals. The frequency of reinforcement in both conditions is governed by a variable-interval schedule and, at first, is the same value in the presence of both stimuli. Then as the frequency of reinforcement associated with one of the stimuli is lowered, the rate of responding in the presence of the other stimulus increases. Conversely, if the frequency of reinforcement is increased in the presence of one of the stimuli, the response rate in the presence of the second shows a decline, although the reinforcement frequency associated with it has remained unchanged throughout.

Similar phenomena have been recorded over the years in various situations. Verplanck (6) and Solomon (5) observed it in running speed in a maze and in jumpstand latency respectively. A contrast effect has occurred very often in studies in which a reinforcement parameter (i.e., magnitude) is changed without an accompanying stimulus change (2). More recently, studies have

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on May 20, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This paper is based on a portion of a thesis submitted in partial fulfillment of the requirements for the Ph.D. degree at Michigan State University, 1963. The author is indebted to M. Ray Denny for directing the research.



noted the facilitative effect of punishment on unpunished responding (1). Holz *et al.* (3) found that changing from a punishment condition to a non-punishment condition repeatedly produced a postpunishment rate that temporally overshoot the prepunishment level; here the effect is called compensation.

The present study attempted to determine if a high-effort response requirement during training has a facilitative effect on low-effort responding during extinction; also, the extent to which this effect can be manipulated. Specifically, the question is, "Does high-effort training augment resistance to extinction of a low-effort response, as compared to low-effort training and low-effort extinction conditions?" Further, if a contrast effect is present, can it be reduced by introducing a small amount of low-effort training after high-effort training and prior to low-effort extinction.

## B. METHOD

### 1. *Subjects*

Thirty male rats, both albino and grey hooded, served as Ss. The animals were approximately 90 to 140 days old at the beginning of the experiment and were from the colony maintained by the psychology department at Michigan State University. Age and strain differences when analyzed had no significant effect upon performance.

### 2. *Apparatus and Procedure*

The apparatus employed was a Skinner box 14 inches long, 14 inches deep, and nine and one-quarter inches wide. The sides were clear plexiglass; while the bar (which was four inches long, three-quarter inch wide and one-eighth inch thick) was of red plexiglass and protruded one inch into the box at a height of three and one-half inches from the floor. A metal food cup was located just to the right of the bar, one and one-eighth inches from the floor. The bar could be counterweighted, so that the effort required to activate the food-delivery mechanism was as low as five grams, as high as 80 grams, or any weight in between. An electric counter recorded all bar-press responses that activated the feeding mechanism. The food reward was a 45-mg pellet (P. J. Noyes).

All Ss were kept on a reduced diet of eight grams of Wayne Lab Blox in their individual home cages for a period of eight to 10 days prior to beginning the experiment. Water was available at all times, and the animals were handled for three-minute periods daily. The Ss were under 21 hours of food de-

privation at the beginning of each session and were fed eight grams of food in their home cages at the end of each session.

At the beginning of training, the bar required only a five-gram force to activate the delivery mechanism. Three food pellets were available in the food cup the first time the animal was put in the chamber. After those pellets were eaten, *E* manually supplied approximately 15 additional pellets, with the click of the feeding mechanism paired with the arrival of food. After the bar-press response was shaped, each *S* made 25 reinforced responses per day until training was completed. Each *S* was extinguished on the day following the completion of training, the dependent measure being the number of responses made to a five-minute no-response criterion.

### 3. Description of Groups

Table 1 summarizes the three groups used. The letter and number designations for each group indicate light (L) or heavy (H) bar loading during training, the number of training reinforcements, and the bar-loading (L or H) during extinction.

*a. Group L 200 L.* All *Ss* made a total of 200 reinforced responses, all of which were on a light bar loading (five grams), and were then extinguished on a five-gram bar.

*b. Group H 120 L.* After the initial shaping, the bar loading in this group was gradually increased until at the end of 80 responses all *Ss* were able to depress an 80-gram bar. This was followed by 120 responses on the 80-gram bar and, then, extinction on a five-gram bar.

*c. Group HL 120 L.* After the initial shaping, the bar loading in this group was gradually increased until at the end of 50 responses all *Ss* were able to depress an 80-gram bar. This was followed by 120 responses on the 80-gram bar. These were followed by 30 responses on a five-gram bar and

TABLE 1  
SUMMARY OF GROUPS

Group	<i>N</i>	Bar loading and number of reinforced responses during training	Bar loading during extinction
L 200 L	10	200 on five grams	Five grams
H 120 L	10	80 on graduated loadings, 120 on 80 grams	Five grams
HL 120 L	10	50 on graduated loadings, 120 on 80 grams, 30 on five grams	Five grams



extinction on a five-gram bar. The 30 interspersed light-bar responses were used in an attempt to minimize the contrast effect.

### C. RESULTS

Table 2 summarizes the data obtained and the comparisons made. The method of analysis consists of separate  $t$  tests of the mean number of responses made during extinction by each group. Two-tailed tests were used in each comparison.

The facilitative effect of heavy-bar training<sup>3</sup> on light-bar extinction is dramatically revealed in comparison No. 1. It can be seen that the H-120-L group made *twice* as many responses during extinction as did the L-200-L group, a difference significant beyond the .0005 level. The magnitude of this difference suggests that single-organism replication can reliably be achieved.

The success of attempting to manipulate the facilitation can be seen in comparison No. 2. The insertion of a small amount of light-bar responding (30 responses) following heavy-bar training significantly reduced light-bar extinction responding on the following day. When compared to Group H 120 L ( $S$ s in which did not receive the 30 light-bar responses), the difference in extinction responding is significant beyond the .005 level.

Comparison No. 3 was an added check on the extent to which the facilitation had been reduced in Group HL 120 L. Since this comparison is nonsignificant, it is possible to conclude that the facilitation was essentially eliminated by the insertion of the 30 light-bar responses prior to extinction.

### D. DISCUSSION

Of interest is the fact that the facilitation observed occurred in the face of the opportunity for generalization decrement. This hypothesis implies that

TABLE 2  
COMPARISONS OF THE MEAN NUMBER OF RESPONSES MADE DURING EXTINCTION

Comparisons	Group	Mean	$S$
1*	L 200 L	155.7	56.2
	H 120 L	310.8	
2**	H 120 L	310.8	57.9
	HL 120 L	193.0	104.0
3***	L 200 L	155.7	56.2
	HL 120 L	193.0	104.0

\*  $t = 6.55$ ,  $df = 18$ ,  $p < .0005$ .

\*\*  $t = 3.25$ ,  $df = 15$  [ $F$  is significant; so  $df$  computed by Welch's formula (7)],  $p < .005$ .

\*\*\*  $t = 1.03$ ,  $df = 14$  [ $F$  is significant; so  $df$  computed by Welch's formula (7)],  $p < .20$ .

extinction should be hastened to a degree that depends upon the magnitude of the differences in stimulation between conditioning and extinction. In the present study, the prediction is that animals trained on a heavy bar and extinguished on a light bar extinguish faster than Ss trained on a light bar and extinguished on a light bar. This, of course, was *not* the case.

Of particular interest is the relationship of the facilitation achieved by varying response effort to areas of behavioral control through aversive stimulation: i.e., punishment. The similarity of the facilitation observed here and the compensation phenomenon noted in punishment studies suggests that other effects of manipulating the effort required for reinforcement might be similar to those of punishment.

#### E. SUMMARY

Three groups of rats were trained to bar press for food on a continuous-reinforcement schedule. Group I was trained to press only a light (5 gm.) bar and was extinguished at the same bar loading to a five-minute no-response criterion. Group II was trained to press a heavy bar (80 grams) and was extinguished on a light bar. This group made over twice as many responses during extinction as did Group I. Ss in Group III were also trained to press a heavy bar, but were given a small amount of light-bar training just prior to extinction on the light bar. This small amount of light-bar training significantly reduced the facilitation in extinction responding noted in Group II.

#### REFERENCES

1. BRETHOWER, D. M., & REYNOLDS, G. S. A facilitative effect of punishment on unpunished behavior. *J. Exper. Anal. Behav.*, 1962, **5**, 191-199.
2. CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467-515.
3. HOLZ, W. C., AZRIN, N. H., & ULRICH, R. E. Punishment of temporally spaced responding. *J. Exper. Anal. Behav.*, 1963, **6**, 115-122.
4. REYNOLDS, G. S. Behavioral contrast. *J. Exper. Anal. Behav.*, 1961, **4**, 57-71.
5. SOLOMON, R. L. Latency of response as a measure of learning in a "single-door" discrimination. *Amer. J. Psychol.*, 1943, **56**, 422-432.
6. VERPLANCK, W. S. The development of a discrimination in a simply locomotor habit. *J. Exper. Psychol.*, 1942, **31**, 441-464.
7. WALKER, H. M., & LEV, J. Statistical Inference. New York: Holt, 1953.

Department of Psychology  
Illinois Wesleyan University  
Bloomington, Illinois





## IDENTIFYING PLACEBO REACTORS\*<sup>1</sup>

*Department of Psychology, Utah State University*

---

HEBER C. SHARP

---

### A. INTRODUCTION

Trouton (9), in a review of research involving placebos and their use, indicated a serious lack of studies concerned with the personality of placebo reactors. Beecher, Keats, Mosteller, and Lasagna (1) pointed out that among their placebo reactors little or no difference was found between percentages of those obtaining relief and those not obtaining relief from the drug being tested. Among nonreactors a significant difference was found. The placebo reactors tend to influence adversely the experimental results. Rosenthal and Frank (7) state that "If we do not control for non-specific factors like the placebo effect, we cannot know whether the effects predicted from a theory lead to or result from improvement based on non-specific effect" (7, p. 299).

Some indications of personality factors identifying placebo reactors is suggested in the study by Lasagna, Mosteller, Von Felsinger, and Beecher (3). They report an investigation of postoperative pain in which all the consistent placebo reactors stated that hospital care was "wonderful," whereas only four out of 16 nonreactors felt the same way. Some Ss consistently responded to placebo and others did not; so the authors suggested the possibility of a personality continuum of this behavior. Differences between reactors and non-reactors were reported in response to vocabulary items, Rorschach plates, anxiety, age, religious activity, and interest in religion. Tibbets and Hawkins (8) reported greater anxiety in placebo reactors than in nonreactors.

Linton and Langs (4) were able to differentiate between reactors and nonreactors, using a 74-item questionnaire. They state that strong placebo reactors are poorly defended and tend to be insensitive, passive, and to have a flattened affect. Those described as weak placebo reactors were sensitive to both external and internal cues, were more intellectually curious, and functioned better intellectually than did strong placebo reactors. Luoto (5) reported that placebos affected differently individuals scoring in the extremes of

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on June 3, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This study was supported in part through a grant from Utah State University Research program.



an extraversion-introversion dimension, whereas extreme groups on neuroticism were much alike: i.e., under placebo influence, high neuroticism influenced introverts and extraverts differently. Question: Is placebo reaction a unitary trait? Lyster *et al.* (6) found that instructions appropriate to presumed drug effect produced a deterioration of performance on simple motor tasks, whereas instructions inappropriate to presumed drug effects counteracted much of the drug-produced decrement.

### B. PURPOSE

In this research, a placebo reactor was defined as that individual who reported as much or more reaction to the placebo pill than was reported for the drug. The concern of this study was two fold: (a) to appraise the frequency of placebo reactors in a sample of a college population and (b) to develop an instrument that would predict the potential reactor. The hypotheses are as follows:

1. There will be no significant number of placebo reactors in the sample of a college population.
2. Low scores on the General Attitude Variability Inventory will *not* indicate nonreactors to a placebo.
3. High scores on the General Attitude Variability Inventory will *not* indicate a potential reactor to a placebo.

### C. METHOD

Students enrolled in the general-psychology sections served as Ss. Two drugs (five mg dextroamphetamine and two grains of caffeine) and a placebo (milk sugar) were used. Drugs and placebo were in identical capsules prepared by a pharmacist. Only the pharmacist knew which capsules contained the drug. E labeled each capsule with the S's name and the order it was to be given. A graduate assistant met the Ss in small groups and administered the capsules; he was not aware of the contents of the capsules. His instructions to the Ss were as follows:

One of the pharmaceutical companies has asked that the university participate in testing a new drug. Local physicians have assured us that the drug will not harm you but is designed to "pick you up." Since we are interested in your reaction to the drug, please fill out one of these questionnaires at the close of the period [Ss studied in the experimental room for one hour] and take a second questionnaire which is to be filled out this evening. Leave the one you have just filled out with me and bring the second one to class tomorrow and give it to your instructors.

Instructors called for the questionnaires at the beginning of class next day.

In the fall of 1963 one section of 88 students was selected as a pilot group. All Ss were given a capsule containing a drug and the placebo capsule; half received the placebo first, half the drug; and five to seven days elapsed between administration periods. Ss completed two questionnaires. Test items concerned anxiety, religion, self-sufficiency, self-concept, dependency, and dominance and were combined into a General Attitude Variability Inventory consisting of 158 items. Each scale was constructed so that a high score contributed to the negative aspect of the trait. All Ss were given this inventory.

During the winter quarter of 1964, two sections of general-psychology students were given the inventory. On the basis of high and low scores 60 subjects (30 females and 30 males) were selected and given both the drug and placebo capsules in a counterbalanced order.

#### D. RESULTS

The pilot group of 88 Ss was given both the drug and placebo capsules in a counterbalanced order. Table 1 summarizes the Ss' reactions.

Eleven Ss did not complete all of the experimental procedures and were dropped from the experiment. Of the remaining Ss, 38.96 per cent reported responses to the placebo pill as great as or greater than those to the drug. Of the males, 33.9 per cent, and of the females, 55.6 per cent indicated a reaction to the placebo. A chi-square test shows a highly significant number of Ss to be placebo reactors ( $p < .001$ ).

The General Aptitude Variability Inventory (GAVI) was administered to the group. Most Ss who reacted to the placebo scored high on the inventory. Females generally scored higher than males. Cutting points were determined, so that high scores indicated a reactor; low scores, a nonreactor. Low scores frequently indicated a subject who showed none or very minor reactions to either the drug or placebo. Of the females who completed the experiment but scored between the cutoff points of 58 and 89, three reacted to the placebo and six did not react to it. Among the male Ss who scored between the cutoff points, 45 and 76, 27 did not react to the placebo and five reacted to it.

TABLE 1  
DISTRIBUTION OF PLACEBO RESPONSES

Population	Male	Female	Total
Reactors	20	10	30
Nonreactors	39	8	47
Dropouts	8	3	11



Table 2 shows the intercorrelations between subtests of the GAVI. The relatively high intercorrelations imply that, with the exception of the religious area, there is considerable overlap in the factors affecting *S*'s response to the items used.

TABLE 2  
INTERCORRELATION AMONG SUBTESTS OF THE GENERAL ATTITUDE VARIABILITY INVENTORY  
(*N* = 77)

Scale	1	2	3	Scale 4	5	6
1. Religion						
2. Anxiety	-.06					
3. Self-sufficiency	-.12	.85				
4. Attitude toward self	-.25	.74	.82			
5. Dominance	.02	.80	.91	.76		
6. Dependency	.00	.57	.69	.63	.66	
7. Total score	.11	.84	.89	.76	.90	.75

A regression analysis of the coefficients indicates that anxiety, self-sufficiency, and dominance seem to be the best predictors for reactor and nonreactor, with dominance contributing most to the prediction.

The second phase of the experiment involved the administration of the GAVI to two sections of elementary general-psychology students. On the basis of high and low scores two groups were drawn (30 males—14 scoring high and 16 scoring low; and 30 females—15 scoring high and 15 scoring low). Two female low scorers withdrew from school prior to completion of the experiment. All *S*s were given both a placebo and a drug in a counter-balanced double-blind design. Chi-square tests indicate that the high scorers on the GAVI reacted significantly to the placebo ( $p < .001$ ), and low scorers did not react to the placebo ( $p < .001$ ). The differences held for the total group as well as for the groups according to sex.

In an early analysis of data from the selected reactor-nonreactor groups, biserial *r*'s were computed for each of the subgroups and for each of the subtests. Table 3 summarizes the results. The complete lack of any relationship between the female reactors and nonreactors to the subtests implies that relationships shown with the total group received the major contribution from the male groups.

On the basis of the correlational data, three subtests seemed to predict better the reactor and nonreactor *S*s. These three areas are anxiety, self-sufficiency, and dominance. Analysis-of-variance tests were employed and *t* tests were used to determine whether or not each subtest discriminated between

TABLE 3  
POINT-BISERIAL CORRELATIONS BETWEEN PLACEBO REACTORS AND NONREACTORS  
ON THE SUBTESTS OF THE GAVI

Scale	Males		Females	
	Reactors ( <i>N</i> = 14)	Nonreactors ( <i>N</i> = 16)	Reactors ( <i>N</i> = 15)	Nonreactors ( <i>N</i> = 13)
Religion	-.13	-.06	.00	.00
Anxiety	.33	-.38	.00	.00
Self-sufficiency	.25	-.15	.00	.00
Attitude toward self	.31	-.18	.00	.00
Dominance	.16	-.21	.00	.00
Dependence	.13	-.34	.00	.00
Total score	.28	-.45	.00	.00

the reactor-nonreactor variables. For the pilot group of 77, *Ss*, none of the analyses yielded a significant *F* ratio.

The analysis of variance using the reactor-nonreactor groups resulted in significant *F* ratios. Table 4 summarizes these data and shows a consistent difference in reactors *vs.* nonreactors on the three subtests of the GAVI.

#### E. DISCUSSION

Literature on persuasion, placebo reaction, and the Hawthorne effect seem to be concerned with similar phenomena. Each report is concerned with the relative effect that particular experimental arrangements have on participating *Ss*. *Ss* react to the experimental situation in a somewhat unpredictable way. Trouton (9) evidenced concern that such phenomenon may distort experimental findings, and Rosenthal and Frank (7) point to the need for appropriate control.

The data on sex differences seems to be similar to that reported by Hovland and Janis (2), who point out "the disparity between male and female sub-

TABLE 4  
*F* RATIOS AND *t* TESTS FOR REACTOR AND NONREACTOR GROUPS  
ON THREE SELECTED SUBTESTS OF THE GAVI

Scale	<i>F</i>	Sex	<i>F</i>
Anxiety	6.45*	Males	3.47*
		Females	2.18**
Self-sufficiency	8.07*	Males	3.67***
		Females	2.01**
Dominance	7.40*	Males	3.53*
		Females	2.12**

\* Significant at the .01 level.

\*\* Significant at the .05 level.

\*\*\* Significant at the .001 level.



jects in the personality correlates of persuasibility . . . as indicating the influence of culturally sex-typed roles . . ." (2, p. 240). Work in this area suggests, according to Hovland and Janis, that since the culturally determined sex role for boys is less definitive in prescribing reactions to persuasive influences than it is for girls, the male is less susceptible to persuasion than the female is. The culture demands of the girl more acquiescence in relation to her social role; hence she is more easily persuaded because she feels less able than a boy to act out feelings of rebellion. Suggestions from high-prestige sources (from peers or an authority figure) are more accepted by the female than by the male.

Ss' reactions to the experimental situation showed a surprisingly large proportion (33.9 per cent) of males responded to the placebo, and over half (55.6 per cent) of the females showed a similar reaction. The large proportion of male Ss reacting to the placebo suggests that many males in our society may also yield to cultural pressure or to sources of authority; hence tend to conform to preconceived behavior patterns. In this study the higher proportion of females responding to the placebo supported the Hovland-Janis hypothesis; yet the relatively large number of male Ss extends a challenge to the thesis.

The correlation coefficients in Table 3 support the differences between the sexes in the consistency of responses to items on the General Aptitude Variability Inventory. No differences between female nonreactors *vs.* reactors was found. The male Ss showed large differences in response to items on five of the six subtests and on the total score. While the differences did not reach significance, the trend showed a consistent difference between responses to the various items by reactor *vs.* nonreactor male Ss.

One must remember that the feminine role is undergoing some change. It is becoming easier for the girl to obtain work and independence from the dominating authority. Present-day culture is more tolerant of girls living away from home than it was a generation or two ago. The male role is more confused with the threat of military service and with the need for more education and specialization, so that the male is often likely to accept advice and counsel: hence, persuasion, placebo reaction, and perhaps the Hawthorne effect may be more prevalent today among young males and less prevalent among young females. Data from the anxiety, self-sufficiency, and dominance scales indicates consistency of responses for both sexes.

From the data concerned with the regression analysis, placebo reactors generally seem to be those individuals who are subject to anxiety, who feel they are insufficient in perhaps most situations, and often seem to be dominated by

other individuals. The nonreactors, on the other hand, were not inclined to suffer from anxiety, felt relatively sufficient in most situations, and tended to dominate others rather than be dominated by them. The nonreactor seems to have personality characteristics that would lead to a rugged-individual type of classification. The reactor is overly concerned about most situations, fearful, and finds it much easier to go along with others (i.e., is inclined to be persuaded into doing that which the other person desires).

An analysis of variance employing the three areas anxiety, self-sufficiency, and dominance was used to determine whether or not there might be a consistent difference in the responses of both male and female Ss to the items in the scales. Significant differences between reactors and nonreactors were found for both sexes on all three scales.

It is possible that more consistency might be found if one were able to appraise more adequately the appropriate area. Further work needs to be done in which larger samples are employed with an extended check for consistency of placebo responses.

There seem to be two pertinent questions for which we do not have the answers. One is concerned with the consistency of placebo reactions: i.e., will the same S, regardless of sex, react consistently to situations of placebo nature? Second, does this personality show a more marked tendency to react positively to persuasion?

Bezcher *et al.* (1) point to the inconsistency of placebo reactors. Lasagna *et al.* (3) refer to consistency of placebo reaction among some reactors and imply that some do not consistently respond to placebo. Linton and Langs (4) refer to strong and weak placebo reactors. Our data suggest some inconsistencies among the placebo reactors, while Lasagna *et al.* (3) suggest the possibility of a personality continuum.

## F. CONCLUSIONS

1. A highly significant number of the sample of college students reacted to the placebo. The percentage of female Ss exceeded that of male Ss.
2. The General Aptitude Variability Inventory differentiated generally between the reactor and the nonreactor, with the areas of anxiety, self-sufficiency, and dominance as the best predictors.
3. High scores on the GAVI indicated a reactor; low scores, a nonreactor. Female Ss tended to score higher than male Ss, and cutting points were higher for female Ss than they were for male Ss.



## REFERENCES

1. BEECHER, H. K., KEATS, A. S., MOSTELLER, F., & LASAGNA, L. The effectiveness of oral analgesics (morphine, codeine, acetylsalicylic acid) and the problem of "placebo reactors" and "nonreactors." *J. Pharmacol. & Exper. Ther.*, 1953, **109**, 393-400.
2. HOVLAND, C. I., & JANIS, I. L. (Eds.), *Personality and Persuasibility*. New Haven: Yale Univ. Press, 1959.
3. LASAGNA, L., MOSTELLER, F., VON FELSINGER, J. M., & BEECHER, H. K. A study of the placebo response. *Amer. J. Med.*, 1954, **16**, 770-779.
4. LINTON, H. B., & LANGS, R. J. Placebo reactions in a study of lysergic acid diethylamide (L.S.D. 25). *Arch. Gen. Psychiat.*, 1962, **6**, 369-383.
5. LUOTO, K. Personality and placebo effects upon timing behavior. *J. Abn. & Soc. Psychol.*, 1964, **68**, 54-61.
6. LYERLY, S. B., ROSS, S., KRUGMAN, A. D., & CLYDE, D. J. Drugs and placebos: The effects of instructions upon performance and mood under amphetamine sulphate and chloral hydrate. *J. Abn. & Soc. Psychol.*, 1964, **68**, 321-327.
7. ROSENTHAL, D., & FRANK, J. D. Psychotherapy and the placebo effect. *Psychol. Bull.*, 1956, **53**, 294-302.
8. TIBBETTS, R. W., & HAWKINS, J. R. The placebo response. *J. Ment. Sci.*, 1956, **102**, 60-66.
9. TROUTON, D. S. Placebos and their psychological effects. *J. Ment. Sci.*, 1947, **103**, 344-354.

*Department of Psychology*  
*Utah State University*  
*Logan, Utah 84321*

## THE RELATIONSHIP OF VERBAL ABILITY AND OTHER COGNITIVE VARIABLES TO THE OPEN-CLOSED COGNITIVE DIMENSION\*<sup>1</sup>

*Department of Psychology, University of Arizona*

SALVATORE V. ZAGONA AND LOUIS A. ZURCHER, JR.

### A. INTRODUCTION

The purposes of this paper are to re-evaluate some of the basic assumptions of Rokeach's theory of dogmatism and to consider possible implications of this re-evaluation for the general concept of the "open" and "closed" mind.

Rokeach formulated the theory of dogmatism as an alternative to the concept of authoritarianism (1) because he felt that authoritarianism, particularly as measured by the F scale, is a measure of *right* authoritarianism and *right* intolerance; whereas dogmatism, as measured by the Dogmatism scale, is a measure not only of *general* authoritarianism and *general* intolerance but also of "the openness-closedness of cognitive systems." In a series of papers (4, 5, 6) and a book (7) summarizing his work in this area, Rokeach describes the empirical studies and theoretical logic that guided him in the formulation of his theory.

The present paper will not attempt to evaluate the Dogmatism scale in terms of its ability to measure general authoritarianism or intolerance; instead, the authors' discussion will be restricted to that aspect of the theory that deals with the dimension of openness or closedness of cognitive systems. Rokeach carefully distinguishes this latter dimension from the more widely understood notion of intellectual rigidity:

Whereas dogmatism refers to total cognitive *organizations* of ideas and beliefs into relatively closed ideological systems, rigidity, when genotypically conceived, refers solely to the degree of isolation between regions or to a "property of a functional boundary which prevents communication between neighboring regions"; when phenotypically conceived, rigidity is defined in terms of the way a person or animal attacks, solves, or learns *specific* tasks and problems. Thus, dogmatism is seen as a higher-order and more complexly organized form of resistance to change.

\* Received in the Editorial Office, Provincetown, Massachusetts, on June 8, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This paper was read at the regional convention of the Rocky Mountain Psychological Association, Salt Lake City, Utah, May 7, 1964.



While dogmatism may well be hypothesized to lead to rigidity in solving specific problems, the converse is not necessarily true (4, p. 196).

More specifically, the distinction between the variables of rigidity and dogmatism may be related to the stages of problem solving. The inability to overcome sets and replace them with new ones is a function of rigidity. The inability to integrate new sets after the old sets have been overcome is a function of dogmatism. This distinction was demonstrated in a study by Rokeach, McGovney, and Denny (8). Subjects scoring high on rigidity manifested significantly greater difficulty than subjects low on rigidity in overcoming the individual sets on the now-famous doodlebug problem. But these two groups did not differ from each other, after the initial sets had been overcome, in their ability to integrate the new beliefs into the problem solution.

One might ask whether differences in intellectual ability may have contributed to the differences shown by the subjects in integrating new beliefs into the problem solution. Rokeach has maintained that dogmatism (specifically, its associated open-closed cognitive dimension) is not influenced by differences in intelligence:

Before going on any further in the interpretation of these findings it would help to know to what extent they can be accounted for by differences in intelligence between the groups. A rough measure of intelligence was available for most of the subjects—the American Council on Education Test. Correlations found between intelligence and scores on the Dogmatism Scale are  $-.02$ , and the correlations are generally small between intelligence and ability to solve the Doodlebug problem (7, pp. 190-191).

Rokeach cites an unpublished thesis by Ehrlich (2) in which a correlation between dogmatism and intelligence (as measured by the Ohio State Psychological Examination) was found to be  $-.01$ .

The support for Rokeach's concept of cognitive openness-closedness rests heavily upon a comparison of the manner in which high and low scorers on the Dogmatism scale are able to integrate new sets into a belief system after old sets have been overcome. It must then be ascertained that the differences noted cannot be attributable to differences in intellectual abilities. The present paper bears upon this latter question, although this was not one of the original purposes of the series of studies being conducted in this and related areas.

#### B. PROCEDURE

Rokeach's Dogmatism scale was administered to 517 freshman students enrolled in the elementary-psychology course at the University of Arizona. The 30 highest scorers and the 30 lowest scorers were selected and assigned to two

special conference sections—the high dogmatics to one section, the low dogmatics to another. (Students in the elementary-psychology course attend two general lectures and one conference section weekly. About 30 students attend each conference section, thus the experimental groups in size were no different from other conference sections.) In the initial design, the authors employed two approaches to the observation and measurement of the experimental variables. First, as instructors for the conference sections, they utilized the classroom to make semester-long observations of the characteristic performance and nature of interaction of the high-dogmatic and low-dogmatic sections respectively. In addition, the experimenters administered in class several paper-and-pencil tests of personality; political, and economic ideology; value orientations; and various other measures. Second, subjects were drawn from the two conference sections for small-group studies in an attempt to determine systematic differences between high-dogmatic and low-dogmatic groups in level of interaction, problem-solving behavior, development of leadership, responses to challenges by authority figures, and other pertinent distinguishing characteristics. (A more detailed description of this initial design and the results obtained is being prepared for presentation in a later report.)

Both experimenters were struck by the differences in general scholastic performance and level of participation and interaction between the two experimental conference sections. Particularly noticeable were the differences in approaches to the subject matter of general psychology, in the capacity and the willingness to communicate, in getting at the "heart" of theoretical issues, and, in general, in the ways in which each group met the intellectual challenges presented by a first course in a social science. In short, the experimenters detected significant intellectual (as well as personality) differences between the two groups: i.e., differences that should not have appeared if much of the empirical support for Rokeach's theoretical formulations is to retain its meaningfulness. These were the observations that led the authors to investigate the relationship between some intellectual variables and Dogmatism scores.

The verbal subtest of the College Qualification Tests was given to all members of the high-dogmatic and low-dogmatic conference sections. In addition, each group was given a 17-item modified remote-associations test, constructed by the present authors from ideas proposed by Mednick (3). The test is designed to measure creativity by requiring the subject to "form associative elements into new combinations by providing mediating connective links" (3, p. 226). Some sample items of the test, with a mediating link in parentheses, follow:



alley	tail	nip	(cat)
tin	can	beer	(can)
burn	bathe	bonnet	(sun)
bell	fly	exams	(bar)

To the authors' knowledge, neither standardization data nor reports concerning the reliability or validity of this measure of creativity are available. The present authors included the test because it appeared to make the kinds of integrative and synthesizing demands almost certain to be influenced by Rokeach's concept of open-closed cognitive dimension. Whether this latter supposition is warranted or whether the remote-associations test is merely another measure of some kind of verbal fluency is yet to be determined.

Finally, following an analysis of the results of the foregoing measures, the Dogmatism scores, College Qualification Verbal Subtest Scores, and raw scores on the first midterm examination in elementary psychology were collected and compared for the entire population of approximately 500 subjects from which the extremes had been drawn.

### C. RESULTS

As has been indicated, the most striking result of this study was in terms of differences in class performance between the high-dogmatic and low-dogmatic groups. Intellectual lethargy characterized the atmosphere of the high-dogmatic classroom. An *unwillingness to relate* seemed to prevail: an unwillingness to relate to the subject matter, to the instructor, to other students. Hesitancy, indeed fear, seemed to accompany even the shortest, most-perfunctory contributions made in response to the instructor's (i.e., experimenter's) urging for broader participation in class discussion.

The atmosphere of the low-dogmatic classroom was in virtually every way precisely the opposite. An air of free discussion pervaded each class meeting; spirited debates of knotty problems and of theoretical issues in psychology were difficult to check once they got started; and the "instructor's dream" of class enthusiasm and general intellectual curiosity was more than realized. (Experimentally organized discussion and debating sessions between the high-dogmatic and low-dogmatic groups yielded even more striking differences. The results of these studies are being prepared for publication.)

For each of the variables for which scores were obtained, differences between the high-dogmatic and low-dogmatic groups were obtained significant beyond the .01 level. These results are shown in Table 1. For the entire group of approximately 500 subjects, correlations among the variables of dog-

TABLE 1  
COMPARISON OF HIGH-DOGMATIC AND LOW-DOGMATIC GROUPS

Variable	High-dogmatic group		Low-dogmatic group		<i>t</i>
	Mean	SD	Mean	SD	
Verbal score	50.1	10.8	58.2	11.8	2.6*
Remote associations	6.8	2.0	8.2	2.5	7.5**
Midsemester examination	42.5	9.6	52.2	10.1	3.5**

\*  $p < .01$ .

\*\*  $p < .001$ .

matism, examination grades and CQT scores, though relatively small, are consistently statistically significant [see Table 2].

We do not offer these data for the purpose of establishing that cognitive open-closedness is, after all, a function of intelligence. No study has yet definitively quantified the extent of the relationship between intelligence (when adequately measured) and the "open" and "closed" mind. But we believe that this study has established for the first time a clear and unmistakable relationship between these variables; i.e., that differences in performance found by Rokeach between high and low dogmatics in solving various integration-type and synthesis-type tasks are not attributable *only* to variables that are relatively independent of intellectual abilities. The data we present demonstrate that despite Rokeach's repeated assertions to the contrary, a statistically significant relationship exists between factors generally associated with intelligence, and scores on the Dogmatism scale. However, our own conviction of the strength of this relationship rests not so much upon a comparison of verbal test scores and the like. It rests much more heavily upon a comparison of two groups of persons observed over a full semester.

In subsequent reports the authors hope to confirm several hypotheses suggested by Rokeach's formulations, particularly with regard to their relevance to the dynamics involved in group functioning and structure. It is unfortunate, therefore, that the observations reported in this paper cause some of the

TABLE 2  
PEARSON PRODUCT-MOMENT INTERCORRELATIONS AMONG MIDSEMESTER EXAMINATION GRADES, VERBAL SCORES, AND DOGMATISM SCORES\*

Variable	Dogmatism scores	Verbal scores
Dogmatism scores	.70**	
Verbal scores	-.18	.43
Examination grades	-.20	

\* All  $r$ 's significant beyond the .001 level.

\*\* Coefficient of stability over a 15-week interval ( $N = 440$ ).



assumptions basic to this important theory to be seriously questioned. If intelligence contributes significantly to dogmatism-scale scores, then the procedures by which Rokeach sought to validate the Dogmatism scale can be questioned. These procedures involved the observance of differential performance of high dogmatics and low dogmatics in problems requiring mainly idea integrations; however, other comparisons were made of flexibility of musical tastes, etc. It is not unreasonable to suppose that differences in these kinds of behavior may be related to differences in intelligence. We have established that a statistically significant difference in intelligence was demonstrated in our study not only by a comparison of extreme groups but over the entire range of scores for over 500 subjects. We concede that (although significance beyond .01 was reached in both cases) the size of the correlation coefficient is small and that the prediction of individual scores from one distribution to another would be exceedingly difficult. But it is only fair to indicate that the degree of relationship found is lower than one could expect to obtain if the ranges of scores on the cognitive variables were not restricted. It is unnecessary, of course, to point out that a restriction of the range of scores depresses the correlation coefficient and renders the degree of covariation of the distributions of scores under consideration to be spuriously low. The higher up the educational ladder that intelligence measures are taken, the more restricted does the range of scores become; and the consequent effect upon the size of  $r$  becomes increasingly significant. As far as is known, this is the first time that the scores of a large sample of first-semester freshmen have been used to relate these variables. Previous studies, including that of Zagona and MacKinnon (9), in which a sample of upper-division students was used, failed to note even a moderately high correlation between the open-closed cognitive dimension and other intellectual variables. It now becomes obvious that this failure was caused by the fact that one of the distributions represented a relatively homogeneous sample.

In commenting upon the size of the correlation between the open-closed dimension and these intellectual variables, it should be noted that the direction and nature of trends can be in themselves meaningful and theoretically significant. In summarizing his work in *The Open and Closed Mind*, Rokeach has commented as follows:

What we are driving at is that in evaluating the outcome of our research effort from a statistical standpoint, the trend of result is a more important consideration than the precise level of significance achieved by a particular set of differences (7, p. 405).

Paradoxically, this concern for trends seems at times to have escaped Rokeach in his own researches, for a persistent trend is evident in a number of his own

studies, demonstrating the direction of the relationship between intelligence and open-closed mindedness, however small this relationship was found to be. For example, in two separate studies, high and low scorers on the Dogmatism scale averaged 5.4 and 5.6 respectively (7, p. 222) and 5.37 and 5.89 respectively (7, p. 277) on the ACE tests. Although these figures are presented as *not* statistically significant, they can be seen to be in the direction of the trend alluded to earlier. In another study (7, p. 262), Rokeach did find a statistically significant difference between intelligence as measured by the ACE tests and Dogmatism scores. On this occasion, he was forced to use samples matched for intelligence in the subsequent experimental procedure. Greater sample size ( $N = 400$ ) and a wider range of scores on the intellectual variable probably contributed to the significance of the noted relationship.

As noted, this paper does not purport to demonstrate the precise relationship that exists between intelligence and the open-closed cognitive dimension. The measures used in this study can be related only indirectly to intelligence. More precise measures of intelligence, as well as a more elaborate procedure of differentiation than can be made by the Dogmatism scale alone, will probably be required if more light is to be shed upon the relationship between cognition and other personality factors.

#### REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. The Authoritarian Personality. New York: Harper, 1950.
2. EHRLICH, H. J. Dogmatism and intellectual change. Unpublished M.A. Thesis. Ohio State University Library, Columbus, Ohio, 1955.
3. MEDNICK, S. The associative basis of the creative process. *Psychol. Rev.*, 1962, **69**, 220-232.
4. ROKEACH, M. The nature and meaning of dogmatism. *Psychol. Rev.*, 1954, **61**, 194-204.
5. ———. On the unity of thought and belief. *J. Personal.*, 1956, **24**, 224-250.
6. ———. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychol. Monog.*, 1956, **70**, No. 18.
7. ———. The Open and Closed Mind: Investigations into the Nature of Belief Systems and Personality Systems. New York: Basic Bks., 1960.
8. ROKEACH, M., MCGOVNEY, W. C., & DENNY, M. R. A distinction between dogmatic and rigid thinking. *J. Abn. & Soc. Psychol.*, 1955, **51**, 87-93.
9. ZAGONA, S. V., & MACKINNON, W. J. Openmindedness and hypotheses of interdependence between libertarian and equalitarian processes. *J. of Psychol.*, 1961, **52**, 347-361.

Department of Psychology  
University of Arizona  
Tucson, Arizona 85721





## AN EMPIRICAL EVALUATION OF THE EFFECT OF ITEM OVERLAP ON FACTORIAL STABILITY\*<sup>1</sup>

*System Development Corporation, Santa Monica, California*

MILES S. ROGERS AND GERALD H. SHURE

### A. INTRODUCTION

Previous investigators have differed on their estimates of the effect of item overlap on test analyses by factor methods. In an early factor-analytic study of the MMPI, Wheeler, Little, and Lehner (15, p. 136) state that "positive item overlap is one way in which the scales are related and should not be interpreted as detracting from empirical relationships based on scores obtained on the scales." On the other hand, Guilford (7) suggests that the existence of item overlap is sufficient reason for refraining from factor analysis of a set of test scales. However, the exact nature of the distortion (if any) produced in the results of such a factor analysis is not evident from the theoretical considerations Guilford raises. Curiously, there are only two empirical studies of the effect of item overlap on factor analysis to be found in the literature. In an early study, Welsh (14) factored truncated scales (all overlapping items removed). Based on the apparent correspondence between their MMPI factor analysis and Welsh's results (14), Kassebaum, Couch, and Slater (9, p. 228) assert that "the practical significance of the problem posed by item overlap may have been exaggerated." This assertion would have been more convincing if the solutions reported in these analyses had been rotated to an objective criterion and if some quantitative index of similarity had been reported. Instead, both these solutions are essentially unrotated centroid analyses, each with one "large" factor highly loaded on most scales and two "smaller" factors.<sup>2</sup> Because most of the test variance in both studies is associ-

\* Received in the Editorial Office, Provincetown, Massachusetts, on June 9, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was supported, in part, by ARPA, under contract SD-97. The authors wish to express their appreciation to Carlos Cuadra, Harry H. Harman, and Burton R. Wolin who commented on an earlier draft of this paper, and to Mary Nickerson for secretarial assistance and editorial suggestions.

<sup>2</sup> Although Kassebaum *et al.* assert that their solution was rotated to "simple structure," comparison of their rotated and unrotated factor loadings shows that only the two less well-defined factors were rotated, and these only slightly. Their large, general factor was unrotated.



ated with a single factor, highly loaded on most scales, it is not surprising that this factor "looks" similar in the two solutions.

In fact, there is little reason to suppose that rotation of these factors to an objective criterion (e.g., varimax) would result in high cross-study similarity. The drastically shortened scales produced by Welsh's truncation method (14) are clearly less reliable than the original scales, and half of the shortened scales have sizeably lower reliability coefficients, averaging .22 (5, p. 474). It would be surprising, indeed, if such markedly altered scales left the rotated factor structure unaltered.

Using a totally different approach, Shure and Rogers (13) report an orthogonal varimax factor analysis of the correlation between MMPI scales produced solely by overlap items. These correlations were obtained by considering the overlap items as the common elements in a common-elements correlation coefficient. Three factors were extracted from the overlap correlations and rotated to normal varimax. One of these was identified with the neurotic triad; another, with the psychotic triad—triad patterns commonly found in MMPI clinical studies. They compared these results with similar factor analyses of the original MMPI scales (1, 3, 9, 15) and Welsh's truncated scales (14). The overlap-factor results were highly similar to factors obtained for the four samples indicated above. In contrast, none of the three factors based on a normal varimax rotation of factors from Welsh's truncated scales (14) resembles these two overlap factors. Differing with Kassebaum *et al.* (9), Shure and Rogers conclude that a significant part of cross-study factor stability may be erroneously introduced or exaggerated by item-overlap artifact rather than being a reflection of stable personality dimensions.

Given the methodological inadequacy of the earlier empirical assessment of item overlap and the potentially serious bias suggested by Shure and Rogers' study, a more adequate empirical appraisal of item overlap was called for. The present study attempted to make a thorough reinvestigation of this problem (*a*) by selecting a test instrument using scales with overlapping items for which cross-study stability of factors had been reported, (*b*) by employing a common, objective-rotation criterion for factoring the correlations among scales, (*c*) by employing an improved technique for eliminating the effect of overlapping items which minimizes alteration in the basic scales, and (*d*) by employing objective criteria for assessing cross-study factor similarity.

The California Psychological Inventory (CPI) was selected for study (6). First, its basic scales are composed of items that overlap on the average 55 per cent with items on one or more other scales, and only two of its 18 scales are entirely free from overlapping items (Communality and Flexibility).

Second, Shure and Rogers (12) previously reported evidence of high factor stability for the CPI. The results of their analysis and several other recent factor analyses of its scales are summarized in Table 1. For each of the factors reported and interpreted in a given investigation, a similar factor with a similar interpretation is found in almost all of the other studies. Using the correlation<sup>3</sup> among factor loadings of the most similar factors in

TABLE 1  
SUMMARY OF CPI FACTOR ANALYSES

Investigators	Population sampled	Rotation criterion	Factor labels
Mitchell and Pierce-Jones (11)	213 female and 45 male teacher trainees	Varimax (orthogonal)	<ol style="list-style-type: none"> <li>1. Adjustment by social conformity (A)</li> <li>2. Social poise or extraversion (B)</li> <li>3. Capacity for independent thought and action (C)</li> <li>4. SuperEgo strength (D or E)</li> </ol>
Crites, Bechtoldt, Goodstein, and Heilbrun (4)	186 female and 186 male stratified into counseling-center clients and nonclient controls	Graphic (oblique <sup>a</sup> simple structure)	<ol style="list-style-type: none"> <li>1. Good Impression<sup>b</sup>—adjustment by adapting self to reality (A)</li> <li>2. Dominance—adjustment by control of external reality (B)</li> <li>3. Communality—test-taking set (E)</li> <li>4. Flexibility—an attitudinal dimension (C)</li> <li>5. Femininity—a trait dimension (D)</li> </ol>
Leton and Walter (10)	114 female and 150 male ninth-grade students	Quartimax (orthogonal)	<ol style="list-style-type: none"> <li>1. A general measure of health and personal efficiency (A)</li> <li>2. Social confidence and drive (B)</li> <li>3. Feminine-masculine dimension (D)</li> <li>4. Social dependence and personal independence (C)</li> <li>5. Attempting to "look good" and impress others (E)</li> </ol>

<sup>a</sup> While the reference-structure matrix of the oblique solution is employed for comparison with the orthogonal factors from the other analyses, these factors are proportional to the primary factors [see Harman (8, pp. 277-280)].

<sup>b</sup> Factor names in this study are associated with a particular reference scale.

<sup>3</sup> The correlation coefficient is not intended to have any geometrical interpretation that might be inferred from the mathematical model used in factor analysis. It is used only to indicate the correspondence between patterns of factor loadings as was done by Burt (2, p. 185) in his use of the "unadjusted correlation" coefficient.



TABLE 1 (*continued*)

Investigators	Population sampled	Rotation criterion	Factor labels
Shure and Rogers (12)	300 male college freshmen and sophomores; high, medium, and low ACE groups ( $N = 100$ each)	Varimax (orthogonal)	1. Personal integrity and mental health (A) 2. Social poise or extraversion (B) 3. Capacity for flexible and independent thought and action (C) 4. Femininity (D) 5. Contented normativism (E)

each study as an index of the similarity of factor loading patterns, Table 2 illustrates the high degree of factorial stability found for several samples where different criteria for rotation (quartimax, varimax, and graphic oblique simple structure) are employed. It supports the conclusion that there appear to be five relatively consistent factors underlying the scales of the CPI.

In order to remove the effect of item overlap, a number of direct strategies

TABLE 2  
CROSS-STUDY COMPARISONS: CORRELATIONS AMONG FACTOR LOADINGS FOR FACTORS WITH SIMILAR INTERPRETATION

Factor study	Mitchell and Pierce-Jones <sup>a</sup>	Crites <i>et al.</i>	Leton and Walter
Crites <i>et al.</i>			
A <sup>b</sup>	.91		
B	.97		
C	.92		
D	.64		
E	.82		
Leton and Walter			
A	.85	.64	
B	.91	.91	
C	.78	.81	
D	.81	.87	
E	.55	.77	
Shure and Rogers <sup>c</sup>			
A	.99	.91	.86
B	.96	.95	.91
C	.93	.97	.83
D	.64	.94	.77
E	.83	.93	.80

<sup>a</sup> Mitchell and Pierce-Jones extracted only four factors [see Table 1]. Their fourth factor appears to be related to Factor D as well as to Factor E and is correlated with both D and E factors from the other studies.

<sup>b</sup> See Table 1 for original factor labels.

<sup>c</sup> Loadings averaged over three stratification groups.

could be employed. (a) One could—as Welsh (14) did for the MMPI—remove the overlapping items from all scales and factor the intercorrelations of the shortened scales. The resulting truncated scales, however, constitute a significantly modified test instrument. Not only would such scales be expected to be less reliable due to their shortened length, but there is no *a priori* certainty that the items of the shortened scales would constitute a representative sample of the items in the full-length scales. (b) In principle, one could employ a more thorough procedure and replace each overlapping item with an alternate form of the item for each additional scale on which it appears. This would require the development of a set of 240 proved alternate forms of the CPI items to cover all instances of item overlap and would increase the test's length by more than 50 per cent. (c) Instead of alternate forms, duplicate items could be employed with the same increase in test length as in (b). In addition, the effect on the test taker of encountering the same item more than once (some items would have to be repeated as many as six times) might cause him to remember or copy previous responses, thus rendering uncertain the independence of the duplicated items. (d) An intermediate solution, the one finally chosen and described in detail later, uses a random subset of overlap items for each subject. This procedure involves statistical analysis only and is not as costly as providing alternate or duplicate forms of each overlapping item. Moreover, it permits each item to contribute to each of the scales on which it is usually scored.

## B. PROCEDURE

The population consisted of 300 male freshmen and sophomores [see Table 1]. This population was divided into high-ability, medium-ability, and low-ability groups. These groups were essentially nonoverlapping in terms of their scores on the ACE Psychological Examination. Their CPI responses were rescored using scale keys generated by an IBM 1401 computer program such that (a) no item appeared on more than one scale for each subject, (b) it was a matter of chance on which scale each "overlapping" item was included for each subject, and (c) individual scale scores for which items were not tallied were prorated so as to make their range equivalent to that of full-length scale scores. In effect, each scale includes its unique items and a randomly selected portion (averaging less than 50 per cent) of its overlapping items. The reduction in scale length is expected to decrease the scale reliability, but less than would occur with complete removal of these items. Furthermore, the integrity of the test is preserved in that each item contributes to each of the scales on which it was originally included for a random



subset of subjects (approximately 50 per cent of the sample). The scores produced by this method clearly contain no built-in association. Any effect of item overlap should be absent from a factor analysis based on these overlap-free scores.

Intercorrelations among the overlap-free CPI scales were computed separately for the three ability groups and were separately factor analyzed (with communalities estimated by the squared multiple correlations). Five factors<sup>4</sup> for each set were extracted and rotated to the normal varimax (orthogonal) criterion. The factors were then compared with the factors reported by Shure and Rogers (12), based on the original scales where overlapping items were not removed.

### C. RESULTS

#### 1. *Comparisons of Factors Based on Original Scales with Those Based on Overlap-Free Scales*

In order to determine the effects of the presence or removal of overlap items of the CPI, two measures were employed for assessing similarity of original to overlap-free factors.

a. *Correlation of factor loadings.* An index of the degree of agreement<sup>5</sup> between a pair of factors may be obtained by intercorrelating their factor loadings. Table 3 shows the correlations found for the most-similar pairs of factors (the factor based on the original scales which correlated most highly with that based on the overlap-free scales) for each ability group. As the table indicates, each of the five original factors appears to have a unique, highly similar counterpart in the overlap-free analysis except for the D factor in the low-ability group (.52).

TABLE 3  
CORRELATIONS<sup>a</sup> OF FACTOR LOADINGS OF ORIGINAL WITH OVERLAP-FREE FACTORS

Original factor	Overlap-free factor	High	Ability group Medium	Low
A	1	.92	.95	.91
B	2	.97	.92	.93
C	5	.96	.94	.93
D	4	.86	.87	.52
E	3	.92	.94	.87

<sup>a</sup> Only correlations for "most similar" factor pairs are reported.

<sup>4</sup> These factor loadings are on file with the American Documentation Institute. Order document No. 8462 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

<sup>5</sup> See Footnote 3.

These data point to the high degree of factor similarity for overlap-free and original scale data. However, shifts in factor loadings for scales other than those thought to be critical in defining the factor can affect the intercorrelation index between factors. As a result, it is possible for this index to be misleading—particularly where the relationship between compared factors is not very high.

*b. Correlation of factor scores.* As a precaution, a related index, the correlation of factor scores, was computed as a second measure of factor similarity and affords a more direct comparison of factors. A factor score can be interpreted as the contribution of a factor to an individual's test score. Ordinarily, factor analysis produces factor loadings for variables. While factor measurements for individuals cannot be obtained directly, statistical means may be employed to obtain estimates of factors in terms of observed scores. A computer program permitted us to estimate factor scores for all five factors for each individual from both the original and overlap-free scale factor analyses. If individuals receive highly similar factor scores for these two analyses, it would indicate that the factors can be interpreted as measuring the same attribute; highly dissimilar factor scores would suggest that the factors being assessed differ.

Correlations between the sets of factor scores shown in Table 4 clearly indicate that each of the five factors produced by the analysis of original scale scores has a counterpart in the overlap-free factor analysis. Only for the low-ability group do Factors D and E appear with questionable similarity (.53 and .77 respectively).

## 2. *Comparison of Stability in Three Ability Samples of Factors Based on Original and Overlap-Free Scale Data*

The previous comparisons have shown the degree to which the original factors are similar to the overlap-free factors. To decide whether or not inter-factor stability is increased as a result of scale item overlap, we may determine whether the similarity among original scale factors found in different samples is greater than the similarity among overlap-free factors and how much the intersample stability is lowered, if at all, by item-overlap removal. The sizeable correspondence in patterns of factor loadings over the three ability samples made it possible to identify one factor in each group with the most similar one in the other groups. Correlations were then calculated between factor loadings for most-similar factors found in the three ability groups. These correlations were made separately for both the original and overlap-free factors. The results are shown in Figure 1. In all but one



TABLE 4  
CORRELATION BETWEEN FACTOR SCORES BASED ON FACTOR ANALYSES OF ORIGINAL AND OVERLAP-FREE CPI SCALES FOR THREE ABILITY GROUPS

Factors from overlap-free CPI scores	A			B			Factors from standard CPI scores						D			E		
	H <sup>a</sup>	M	L	H	M	L	L	H	M	L	C	L	H	M	L	H	M	L
Factor 1 High Medium Low	.93	.95	.89	.08	.06	.15	-.11	-.07	-.14	-.24	-.09	-.23	-.25	-.04				
Factor 2 High Medium Low	.01	.04	.00	.96	.92	.88	.00	-.14	-.04	.07	-.09	.05	-.08	-.09				
Factor 5 High Medium Low	.22	.15	-.12	.03	.24	.04	.94	.85	.91	.05	.00	-.20	.21	-.22				
Factor 4 High Medium Low	.32	-.25	-.08	.09	-.21	-.35	.06	.08	.18	.86	.88	-.03	-.03	-.04				
Factor 3 High Medium Low	-.09	.15	-.35	.09	.09	-.06	.03	-.01	.02	.02	.16	.90	.82	.77				

<sup>a</sup> H = high, M = medium, and L = low ACE group (N = 100 each).

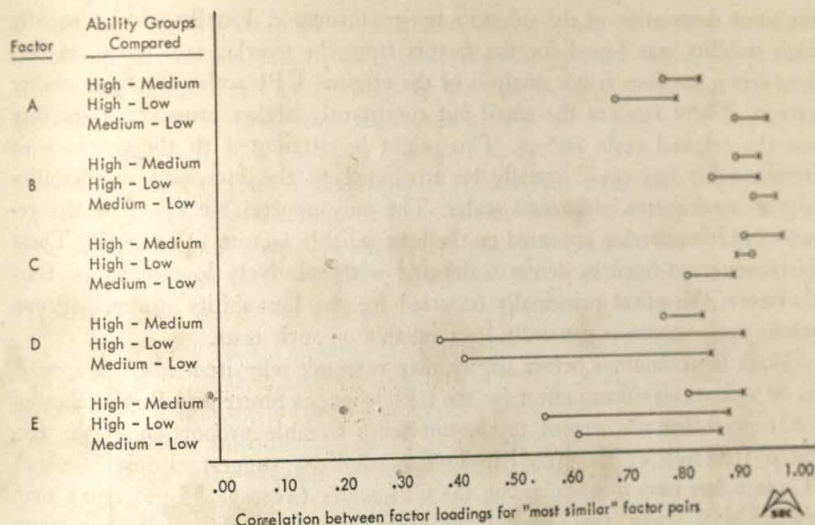


FIGURE 1

CROSS-SAMPLE COMPARISON OF CORRELATIONS AMONG FACTORS BASED ON ORIGINAL SCALES (X) AND CORRELATIONS AMONG FACTORS BASED ON OVERLAP-FREE SCALES (O) FOR THREE ABILITY LEVELS

of 15 comparisons, the correlation between similar overlap-free factors is smaller than the corresponding correlation between original factors.

It should be noted, however, that while overlap-free factors would be interpreted as generally less stable over different samples than are the original factors, the intercorrelations among the former are still relatively high, with 10 of the correlations between the factors falling between .75 and .95. Factors D and E show the greatest reductions in apparent similarity associated with removal of item overlap, particularly in comparisons involving the low-ability group.

#### D. DISCUSSION

Removing item overlap by the method employed in this study appears to have minimal effect on the overall factor structure of the CPI. For each of the five factors found in the analysis of the original scales, a factor was found in the analysis of overlap-free scores with a highly similar, highly correlated pattern of factor loadings. In addition, comparison of factor scores based on the factor analysis of the original and overlap-free scales indicates that either factor solution (with or without item overlap) produces nearly



the same description of the subject's test performance. Finally, almost equally high stability was found for the factors from the overlap-free scores as was found in a previous factor analysis of the original CPI scales for three ability groups. There remains the small but consistently higher cross-study stability for the original scale factors. This might be attributed to the presence of item overlap but could equally be attributed to the increased unreliability of the overlap-free, shortened scales. The only appreciable effect of the removal of item overlap appeared on the least reliable factors (D and E). These factors were defined by fewer scales and with relatively low loadings. Furthermore, this effect principally occurred for the low-ability group, a group whose performance is generally least reliable on such tests.

With these findings before us, we may now ask why item overlap appears to be without significant effect for the CPI when, as Shure and Rogers showed (13), it potentially seemed to account for a sizeable proportion of the factorial structure of the MMPI. Two possibilities suggest themselves: (a) There is less item overlap among the CPI scales (average 55 per cent) than the MMPI scales (average 69 per cent); however, this difference seems small compared with the sizeable degree of overlap in both. (b) The item overlap may not be configured on the CPI as it is in the MMPI. Factor analysis of the latter, it will be recalled, produced a factor structure similar to the results for the factor analysis of the original scales and raised a question of the authenticity of reported MMPI factors. Indeed, the CPI overlap items may not exhibit a factor structure bearing any similarity to the five empirical CPI factors.

In order to evaluate the latter possibility, the CPI interscale common-element correlations (produced solely by item overlap) were factor analyzed. Four factors<sup>6</sup> were extracted and rotated to normal varimax. Table 5 shows

TABLE 5  
CORRELATIONS OF OVERLAP FACTOR LOADINGS WITH THOSE FOR A AND B FACTORS  
FOUND IN SIX SAMPLES

Factor study	Factor B with Overlap I	Factor A with: Overlap II	Overlap III	Factor A with <sup>a</sup> : II and III
Crites <i>et al.</i> (4)	.88	.81	.04	.82
Leton and Walter (10)	.81	.27	.04	.28
Mitchell <i>et al.</i> (11)	.93	.63	.37	.78
Shure and Rogers (12)				
High ACE	.90	.58	.45	.79
Medium ACE	.88	.64	.39	.80
Low ACE	.85	.74	.20	.80

<sup>a</sup> Multiple correlations.

<sup>6</sup> See Footnote 4.

the correlations for six different samples between factor loadings for three of the overlap factors and the previously identified A and B factors. Correlations with the C, D, and E factors are considerably lower and are not presented. These results show that Overlap Factor I is strikingly similar to the set of B factors (average  $r = .88$ ) found in six separate instances. In addition, two other overlap factors (II and III) both appear related to the set of A factors. Since both of these overlap factors correlate most highly with the empirical A factors, a multiple correlation between the two overlap factors and each A factor is also presented. The fourth overlap factor was not found to be related to any of the previously identified factors.

Thus, as in the MMPI, two overlap factors are found which are related to the two most well-defined CPI factors. The fact that these factors were least altered by removal of item overlap diminishes Shure and Rogers' concern over the possible introduction of an erroneous or biasing factor structure based solely on item overlap. Why the CPI overlap-free factors fail to be modified by the removal of the overlap factor structure remains a dilemma. A possible, albeit speculative, explanation which is consistent with our findings is offered.

When the CPI scales were constructed, scales were composed of items that best discriminated among the members of groups thought to be characterized by each scale. As a result of this method, an overlap item became included in more than one scale by virtue of the fact that it discriminated well among members of different criterion groups. By this process of item selection, an item (whether overlap or singly keyed) is likely to represent or reflect all of the scale dimensions introduced by this procedure. If some of these scales are then found to have an even more fundamental underlying factorial unity, the common factor involved would be as likely to contribute to the variance of the overlapping as to the nonoverlapping scale items. As a result, those scales that share a common-factor dimension would be more likely to have overlapping items in common than those that are factorially independent of each other. In this way, the configuration of scale intercorrelations based exclusively on overlapping items may reflect the same factors found in the factor analysis of the scales containing no item overlap. If this hypothesis is correct, the overlap items (no more nor less than the other items) express their common dimensionality. Thus, the factor structure caused by the built-in item overlap correlations would be highly similar to the empirical factor structure of the scales, and removing the effect of item overlap should not alter this factor structure.

Regardless of the validity of the foregoing speculation, the sum of the



present evidence indicates that while cross-study factorial stability of the CPI may be slightly exaggerated by item overlap, the empirical factors are not produced or distorted by the configurations of overlapping items. The present study does not answer the question as to whether or not the procedures for eliminating item overlap from the CPI would yield similar results for the MMPI. Obviously, they would do less violence to the basic scales than Welsh's truncation method.

### E. SUMMARY AND CONCLUSIONS

1. Although strong warnings have been voiced against using factor-analytic procedures with scales employing item overlap, investigators continue to factor such scales. Two empirical studies in the literature yield contradictory implications. One suggests item overlap is of little practical significance; the other suggests that it may be a potential source contributing to erroneous or exaggerated factor stability. The present study attempted to make a thorough empirical reinvestigation of this problem.

2. This study investigated the effect of removing item overlap on the factor composition of the CPI (a test with many overlapping items and reported high cross-study factorial stability). Overlap-free scale scores were obtained by a method of random assignment of overlap items for three samples, each composed of 100 male college freshmen and sophomores. Factor analyses of the interscale correlations based on these overlap-free scores were made for each of the three samples and compared with previously published factor analyses of original scale scores for the same three samples. In both cases, five factors were extracted and rotated to an orthogonal varimax criterion.

3. For each of the three populations, two objective measures of factor similarity (intercorrelation of factor loadings and intercorrelation of derived individual factor scores) showed that each factor in the overlap-free analysis had a highly similar counterpart in the original analysis. Furthermore, the high intersample factorial stability reported in an earlier study was only slightly lower in the overlap-free analysis. The only differences between the two factor solutions seemed most likely to be attributable to the lowered scale reliability associated with the shortened overlap-free scale scores.

4. In addition, the CPI interscale common-element correlations (produced solely by item overlap) were factor analyzed to assess the extent to which there is a built-in factor structure in the configuration of overlapping items. Two overlap factors were found which were similar to the two best-defined

original and overlap-free CPI factors. An interpretation, based on CPI scale construction procedures, is presented to account for this similarity.

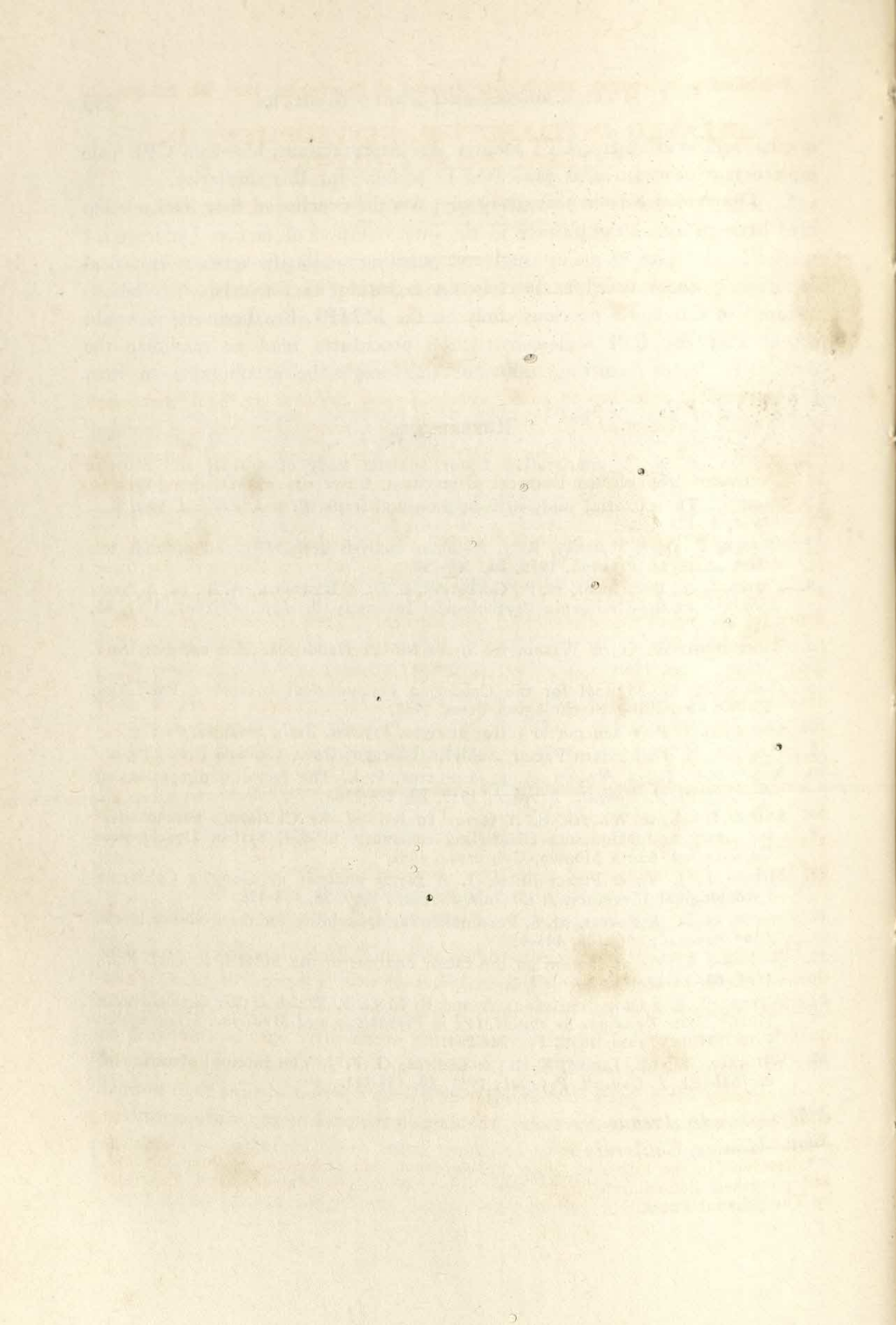
5. The evidence from this study supports the conclusion that item overlap is of little practical significance in the interpretation of factors (at least for the CPI), in spite of an unusual and puzzling similarity between empirical factors and those based exclusively on factoring item overlap correlations obtained in this and a previous study on the MMPI. Furthermore, it would appear that the CPI scale-construction procedures tend to minimize the potentially factor-distorting influence that might be attributable to item overlap.

#### REFERENCES

1. ABRAMS, E. N. A comparative factor analytic study of normal and neurotic veterans. Unpublished Doctoral dissertation, University of Michigan, 1949.
2. BURT, C. The factorial study of temperamental traits. *Brit. J. Psychol. Stat. Sec.*, 1948, **1**, 178-203.
3. COTTLE, E. B., & WHERRY, R. J. A factor analysis of MMPI and aptitude test data. *J. Appl. Psychol.*, 1950, **34**, 260-266.
4. CRITES, J. O., BECHTOLDT, H. P., GOODSTEIN, L. D., & HEILBRUN, A. B., JR. A factor analysis of the California Psychological Inventory. *J. Appl. Psychol.*, 1961, **45**, 408-414.
5. DAHLSTROM, W. G., & WELSH, G. S. An MMPI Handbook. Minneapolis: Univ. Minn. Press, 1960.
6. GOUGH, H. G. Manual for the California Psychological Inventory. Palo Alto, Calif.: Consulting Psychologists Press, 1957.
7. GUILFORD, J. P. When not to factor analyze. *Psychol. Bull.*, 1952, **49**, 26-37.
8. HARMAN, H. H. Modern Factor Analysis. Chicago: Univ. Chicago Press, 1960.
9. KASSEBAUM, G. G., COUCH, A. S., & SLATER, P. E. The factorial dimensions of the MMPI. *J. Consult. Psychol.*, 1959, **23**, 226-236.
10. LETON, D. A., & WALTER, S. A factor analysis of the California Psychological Inventory and Minnesota Counseling Inventory. SP-204, System Development Corporation, Santa Monica, California, 1961.
11. MITCHELL, J. V., & PIERCE-JONES, J. A factor analysis of Gough's California Psychological Inventory. *J. Consult. Psychol.*, 1960, **24**, 453-456.
12. SHURE, G. H., & ROGERS, M. S. Personality factor stability for three ability levels. *J. of Psychol.*, 1963, **55**, 445-456.
13. ———. A note of caution on the factor analysis of the MMPI. *Psychol. Bull.*, 1964, **63**, 14-18.
14. WELSH, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.) *Basic Readings on the MMPI in Psychology and Medicine*. Minneapolis: Univ. Minn. Press, 1956. Pp. 264-281.
15. WHEELER, W. M., LITTLE, K. B., & LEHNER, G. F. J. The internal structure of the MMPI. *J. Consult. Psychol.*, 1951, **15**, 134-141.

2500 Colorado Avenue  
Santa Monica, California





## DELAYED INTELLECTUAL DEVELOPMENT IN CEREBRAL PALSY CHILDREN\*

*Department Rehabilitation Medicine, State University of New York,  
Downstate Medical Center*

---

SAMUEL HIRSCHENFANG AND JOSEPH G. BENTON

---

### A. PROBLEM

The periodic retesting of young cerebral-palsy children at the Department of Physical Medicine and Rehabilitation, Kings County Hospital, has revealed that many of the children show a marked increase in intellectual functioning as they mature. These findings amplify the concept of delayed development by Sarason (2), who has reported on a study by Doll, Phelps, and Melcher (1), of 11 birth-injured subjects between the ages of 7 and 39. The average *IQ* of the subjects below 15 years of age was 51, whereas the mean for those above 15 years of age was 79. The authors considered these findings as evidence for the belief that the mental development of such subjects may continue to a later age than is the case with ordinary mentally retarded subjects and with average normal subjects. Further evidence for the concept of delayed development is presented on the basis of the mental-growth curves of eight subjects who had at least three Binet examinations. In a majority of these cases, there was a noticeable tendency for mental-age growth to continue over a longer period of time than the 14 years generally assumed to be the limit for both normal and retarded persons. The appearance of these cases in such a small sample is considered significant in view of the fact that only five per cent of institutionalized defectives show such increase.

### B. PURPOSE

This paper presents four case studies of young cerebral-palsy children who showed marked increase in *IQ* when retested over a period of time. These findings indicate that the intellectual capacity of cerebral-palsy children frequently is not revealed in their functioning at an early age, as compared to that of normal children. Thus, in addition to the mental-age growth of cerebral-palsy individuals continuing over a longer period of time than normal children (as reported by Sarason), the data in the present paper are consistent

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on June 11, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



with the hypothesis that the rate of intellectual growth of cerebral-palsy children is different from that of normal children.

### 1. Case D. C.

This boy, with a diagnosis of spastic paraplegic, was first seen when he was 3:6. On the Stanford Binet he was successful on only two items at the two-year level. No quantitative measure of his functioning could be determined from the Binet. He was unable to speak, and his social age (as measured by the Vineland Social Maturity Scale) was two years. On the Gesell Developmental Schedules, his motor development was two years; language, 40 weeks; adaptive behavior, 18 months; and personal-social behavior, 15 months.

At the age of 5:4, D. C. achieved a mental age of three years and a Binet *IQ* of 56. His vocabulary level, as measured by the Binet, was 2 years; and he was successful on one task at the 4-year level. At 7 years, D. C. achieved a mental age of 5:4 and a Binet *IQ* of 76. His vocabulary level was 6 years, and he was successful on one task at the 7-year level. Despite the presence of difficulties in concentration and in visual-motor areas, D. C. has a good concept of numbers and is able to read. The prognosis for future intellectual development is much more favorable than when D. C. was evaluated at 3:4.

### 2. Case D. S.

This boy is the first-born sibling of identical twins who were two months premature. At 2:7, formal testing was not possible; but the Vineland Social Maturity Scale indicated a social age of about 16 months. It was concluded that D. S. was a mentally retarded child of whom not too much could be expected in terms of formal intellectual education. At 3:8, he achieved a mental age of 2:3 and a Binet *IQ* of 61. His vocabulary level was found to be 2:6, and his highest level of achievement was at the 3-year level.

At 4:2, D. S. achieved a mental age of 2:10 and an *IQ* of 68. Vocabulary level, the highest level of achievement, was 3 years.

At 6:2, D. S. achieved a mental age of 6:1 and an *IQ* of 99. Vocabulary level was 6 years, and his highest level of achievement was at year XI, at which level he was successful on sentence memory. The early prognosis of mental deficiency was changed to that of average intelligence.

### 3. Case S. D.

This mildly involved child, with a diagnosis of tremor athetoid, achieved a mental age of 2:2 and a Binet *IQ* of 81 when she was first evaluated at 2:8.

Vocabulary level was 2 years, and she was successful on two items at the 2:6 level. She was reported to be functioning at a dull-normal level of intelligence.

At 6:10, S. D. achieved a mental age of 6:8 and an *IQ* of 98. Vocabulary level was 6 years, and her highest level of achievement was on a memory item at the 8-year level. She was functioning at a normal intellectual level and keeping pace with her regular school-grade placement.

#### 4. Case T. B.

This girl achieved a mental age of 2:6 and an *IQ* of 47 when she was first evaluated at 5:4. Her vocabulary level was 3 years, and her highest level of achievement was at 3 years. At the age of 6:2, she achieved a mental age of 3:9 and an *IQ* of 61. This represented an increase of 15 months in mental age over a period of eight months. Vocabulary was 4 years, and she was successful on a definition item at the 5-year level. Although still functioning at a defective level, the improvement in functioning suggests a more favorable prognosis than earlier as regards future intellectual development.

#### C. DISCUSSION

In the repeat evaluations of cerebral-palsy children at Kings County Hospital, such marked increases in *IQ* as in the cases presented are not always encountered. The *IQ*s of some of the children remains constant or decreases. However, increase in the level of intellectual functioning is more the rule than the exception. Of 15 children tested and retested over periods ranging from six months to four years and two months, 11 showed an increase in *IQ* of five or more points. The mean *IQ* of the group at initial testing was 67.7, and the mean *IQ* on the second testing was 76. The difference between the two means is significant at the .02 level. These findings imply or are consistent with the idea that there is a process of delayed development operating in the intellectual growth of cerebral-palsy children. The potential intellectual abilities and resources of cerebral-palsied children frequently are not revealed when such children are evaluated at an early age. On the basis of these findings, the clinician working with cerebral-palsy children should be cautious in his prognosis of the future intellectual development of such children.

#### D. SUMMARY

In the present paper, the authors point out the need for developmental or growth curves of cerebral-palsy children because intellectual development of such children frequently proceeds at a different rate from that of physically



normal children. In future research, effort should be made to isolate those factors that might differentiate the cerebral-palsy child whose development is delayed and the child whose intellectual functioning proceeds at a normal rate.

#### REFERENCES

1. DOLL, E. A., PHELPS, W. M., & MELCHER, R. T. *Mental Deficiency Due to Birth Injuries*. New York: Macmillan, 1932.
2. SARASON, S. B. *Psychological Problems in Mental Deficiency*. New York: Harper, 1949.

*Kings County Hospital Center*  
*451 Clarkson Avenue*  
*Brooklyn 3, New York*

## SELECTIVE TRACE AROUSAL\*<sup>1</sup>

*Department of Psychology, Yeshiva University*

GERALD TOLCHIN AND JOHN CERASO

### A. INTRODUCTION

Interference theory has been a successful framework for organizing many of the laboratory determined facts of acquisition and retention (1). The assumption that this theory can be applied to extralaboratory situations has been widely held, but the precise manner in which interference works in everyday life has not been given the attention it deserves. Recently, Underwood and Postman (5) have attempted to deal with this problem by stating that words in everyday use acquire associative connections with other words. In attempting to learn something new, which involves the pairing of words in new contexts, this new learning may be difficult because it involves competition from old associations.

The principles determining whether a particular old association will interfere with currently learned material are given by the laws of transfer. Chief among these are the similarity and the strength of association laws: The more similar are the stimuli of new and old associations and the stronger the prior association, the greater will be the interference effect; thus, in any particular situation, one should be able to predict the difficulty of new learning if he knows the strength of the old associations and their relative similarity to the new learning.

From everyday experience, however, one is aware that the learning of new verbal material is not accompanied by an elicitation of all or even a small part of the previous associations to the words involved. For example, in learning a set of material that includes the stimulus word "bull," such associates as "pen," "frog," and "dog" probably would rarely occur. On the other hand, if the new learning were to involve one of the old associations (e.g., "bull-frog"), then one would expect learning to be easy. Although the old association is potentially available, it becomes manifest only under certain conditions.

In the example given, why is "bull-frog" so easy to learn even though when

\* Received in the Editorial Office, Provincetown, Massachusetts, on June 11, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This work was supported by a USPHS Predoctoral fellowship (13073) to the first author and a NSF grant (19538) to the second.



"bull-house" (let us say) is presented for learning, the response "frog" does not come to mind? One answer would be that the mere presentation of the pair "bull-frog" in the learning list is sufficient to *arouse* the memory trace of "bull-frog," so that it can be utilized. If "bull-frog" is not a member of the list to be learned, then the corresponding memory trace would *not* be aroused, and it would then be unlikely that "bull" would arouse the "frog" associate. These considerations imply that in transfer in everyday life (in addition to the classical transfer principles) there are factors, such as the arousal by presentation of an old association, that influence new learning.

It is the purpose of the present paper to demonstrate this arousal factor in a transfer design using prepotent associations that can be assumed to be in each subject's repertoire.

## B. METHOD

### 1. *Material*

For the present study, material was selected that would be susceptible to transfer effects from past (pre-experimental) training. This material was derived from a set of compound nouns selected from the Thorndike-Lorge word count (4). The basic idea was to expose briefly a compound noun and then to test the effect of this exposure on the learning of a transfer list. Ten compound nouns were selected, and two paired-associates lists were derived to conform to the A-B<sub>r</sub> paradigm used in conventional transfer studies. These lists were constructed by splitting each compound noun into its two component parts and then rejoining the words, so that the first word of each compound noun was coupled with the second half of a different compound noun. Table 1 presents both the original compound nouns and the two derived lists.

TABLE 1  
ORIGINAL COMPOUND NOUNS AND DERIVED LISTS

Original compound nouns	Derived lists	
	List I	List II
backbone	back-worm	back-coal
barnyard	barn-road	barn-well
bathroom	bath-some	bath-bone
charcoal	char-ball	char-road
farewell	fare-book	fare-some
football	foot-coal	foot-yard
handsome	hand-well	hand-room
notebook	note-yard	note-ball
railroad	rail-bone	rail-worm
tapeworm	tape-room	tape-book

## 2. Procedure

List I was used as the transfer list, and it was learned under three different conditions. Ss in Group A were shown the list of original compound nouns for 30 seconds, under instructions to familiarize themselves with the nouns but not to try to learn the nouns as they would not be tested on this material. Immediately after the 30-second exposure, Ss in Group A were asked to learn the nouns in List I. Ss in Group B followed the same procedure except that they were presented with List II (instead of the original compound nouns) for 30 seconds prior to the learning of the nouns in List I. Ss in Group C learned the nouns in List I without any prior exposure. For all three groups, standard paired-associates instructions were given; and learning continued for 10 anticipation trials. The list was exposed on a memory drum at the rate of two seconds for the stimulus and two seconds for the stimulus and response. Learning was by the method of anticipation recall.

## C. SUBJECTS

Sixty-four undergraduate students at New Paltz State College served as subjects. All subjects were volunteers and naive to verbal-learning experiments. The subjects were assigned randomly to one of the three groups: 14 in each of Groups A and B and 36 in Group C. Additional subjects were obtained in Group C since this group was also used as a control group for another experiment.

## D. RESULTS

Analyses of variance and *t* tests were performed, using the mean number of correct anticipations for each of three criterion measures: (a) the number of correct anticipations over the first two trials, (b) the number of correct anticipations on the 10th trial, and (c) the total cumulative correct anticipations. The data are presented in Table 2. The results are the same regardless of the criterion measure used. Ss in Group A (i.e., the arousal group) took significantly longer to learn the nouns in the transfer list (List I) than did

TABLE 2  
MEAN NUMBER OF ANTICIPATIONS FOR GROUPS A, B, AND C

Measure	Group A ( <i>N</i> = 14)	Group B ( <i>N</i> = 14)	Group C ( <i>N</i> = 36)
Number correct over first two trials	1.00	3.14	4.11
Number correct on 10th trial	4.14	6.71	6.53
Total correct cumulatively	23.36	43.21	45.22



either of the other groups. Further, there were no significant differences between the performances of subjects in Groups B and C.

### E. DISCUSSION

Selective-trace arousal is clearly demonstrated in this experiment. The group given a set for the original compound nouns suffered significantly more interference than did either of the other two groups. One might argue that the 30-second exposure to the compound nouns prior to the learning of the transfer list was sufficient to produce enough learning to produce negative transfer; but, if this were the case, interference should have been demonstrable in Group B. However, there were no significant differences between this group and the group given no pre-exposure, indicating that arousal and *not* learning was the factor involved in producing interference. The results, therefore, are interpreted as meaning that the negative transfer demonstrated by the arousal group was produced by the old associations that had been selectively aroused, and that the control group suffered little from these prepotent associations.

The arousal effect can be viewed as similar to the phenomenon of priming (2, 3). Priming studies show that the presentation of an *item* increases the probability of occurrence of that *item* in an association test. The present study demonstrates that the presentation of a pair of previously *associated items* increases the probability of that *association* occurring in new learning.

The major implication of the present study is that stimuli that have multiple associations do not automatically elicit those associations in new learning and that there are discoverable factors that determine which associates will be elicited: for example, in the typical laboratory study of transfer it could be the fact that the original and transfer lists are learned under very similar conditions (same room, same experimenter, etc.) that leads to reliable arousal of first-list associations during second-list learning. It is probable that the technique employed in the present experiment is only one of a number of methods by which old associations can be selectively potentiated.

### F. SUMMARY

A and B may be associated; yet presentation of A may not elicit B. However, if S were asked to learn A-B he would find it easy. The argument is that presentation of A-B in the learning task arouses an A'-B' memory trace, so that presentation of A reliably elicits B rather than any other (or no) associate.

Three groups of Ss learned the same A-B list of 10 paired words. One group was first briefly shown familiar compound nouns that when rearranged formed the A-B learning list. A second group was briefly exposed to unfamiliar compound nouns that also formed the A-B list. A third group had no prior exposure. The first group had more difficulty in learning A-B than the other groups which did not differ from each other. The brief exposure of the compound nouns made them effective in producing negative transfer.

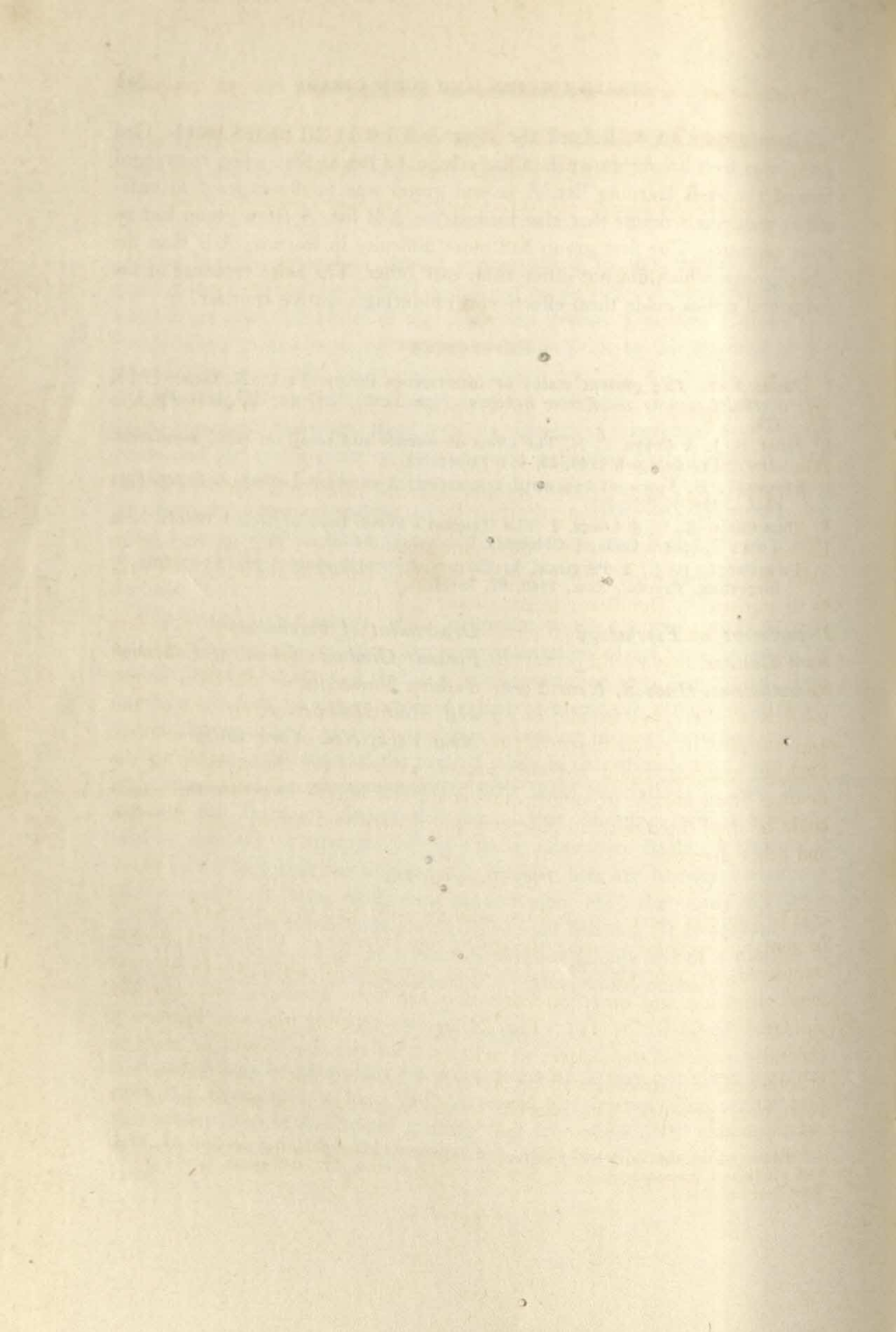
## REFERENCES

1. POSTMAN, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal Learning and Verbal Behavior*. New York: McGraw-Hill, 1961. Pp. 152-179.
2. SEGAL, S. J., & COFER, C. N. The effect of recency and recall on word association. *Amer. Psychologist*, 1960, **15**, 451 (abstract).
3. STORMS, L. H. Apparent backward association: A situational effect. *J. Exper. Psychol.*, 1958, **55**, 390-395.
4. THORNDIKE, E. L., & LORGE, I. *The Teacher's Word Book of 30,000 Words*. New York: Teachers College, Columbia University, 1944.
5. UNDERWOOD, B. J., & POSTMAN, L. Extraexperimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73-95.

Department of Psychology  
Bard College  
Annandale-on-Hudson, New York

Department of Psychology  
Ferkauf Graduate School of Education  
Yeshiva University  
110 West 57th Street  
New York, New York 10019





## SOME CORRELATES OF INNER DIRECTEDNESS AND OTHER DIRECTEDNESS\*

*University of Kansas and State University of New York at Cortland*

---

ANTHONY J. SMITH AND JOHN N. PARK

---

### A. INTRODUCTION

David Riesman's provocative speculations concerning social character (9, 10) have inspired a variety of psychological studies ranging from the nature of disciplinary procedures used in socializing children (2) to the recognition threshold for value-related words (7).

Position on an inner-directed-*vs.*-other-directed continuum has been related to personal data or demographic variables by Kassarian (5), Centers (3), Kallen (4), and Peterson (8). These investigators report findings that are in some degree consistent with Riesman's typology either in the form of trends or of statistically significant relationships.

In the preliminary and final phases of a study by Park and Smith (7), data were collected that would permit the further testing of some of these findings: in particular, those that indicate a tendency toward other direction among subjects who are (a) female, (b) young, (c) from large urban centers, and (d) enrolled in specialties involving an interpersonal or service orientation. Such an analysis would not only provide evidence for the stability of the findings from sample to sample, but would also indicate whether the results could be duplicated with an independently derived measure of inner direction and other direction.

### B. PROCEDURE

The subjects were 226 students who were enrolled in sections of a course in general psychology during the spring and summer of 1958. Each student completed an abbreviated version of a questionnaire designed to measure inner direction and other direction that had been developed and carefully validated by Bell (1, 11). The 26 items composing the test instrument generally required the subject to indicate what response should be made in hypothetical situations involving a conflict between one's own values and peer-group pressure. The possible scores ranged from a minimum of 53

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on June 21, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



(extreme inner direction) to a maximum of 128 (extreme other direction). The obtained scores ranged from 58 to 108 with a mean of 89.9.

In the summer of 1964 records in the Registrar's Office were examined to determine which students had completed work for a degree, the area in which the degree was earned, age, and home town. The population of the home town was then obtained from current census records.

### C. RESULTS

#### 1. Sex

Mean scores on Bell's questionnaire were determined separately for the 115 women and 111 men who composed the total sample. The mean for females was 91.74; that for males, 88.03. The difference was tested by means of the  $t$  test and found to be highly significant ( $t = 3.67$ ;  $p < .001$ ).

#### 2. Age

Following the categorizing method used by Kassarian, the 213 subjects for whom the information was available were classified into three groups with respect to age: 18 and under, 19 to 21, and 22 and above. The mean scores for the three groups were 90.89, 89.90, and 90.40 respectively. A simple analysis of variance reveals no significant differences among the three groups ( $F < 1$ ).

#### 3. Home Town

The 212 subjects for whom the data were available were categorized in six groups on the basis of population of home town: less than 2,500; 2,500 to 4,999; 5,000 to 19,999; 20,000 to 99,999; 100,000 to 399,999; and 400,000 and above. The mean scores on inner direction and other direction for the six groups were 92.80, 90.70, 89.39, 89.85, 88.25, and 89.00 respectively. The data were analyzed by means of a simple analysis of variance, and the means were found to be homogeneous ( $F = 1.53$ ;  $p > .05$ ).

#### 4. Area of Specialization

An examination of the Registrar's records revealed that 130 of the subjects had earned degrees before July, 1964. The areas in which the degrees were earned were determined and were assigned to one of nine major categories. Two categories were dropped because of inadequate sample size (journalism and fine arts). Mean questionnaire scores were determined for each of the remaining categories with the following results: (a) natural science, 83.44; (b) engineering, 85.20; (c) social science, 85.44; (d) human-

ities, 87.88; (e) business, 89.40; (f) education, 91.31; and (g) ancillary medical specialties (occupational therapy and physical therapy), 95.25. The data were analyzed by means of a simple analysis of variance, and the variation among means was found to be significant ( $F = 2.94$ ;  $p = .01$ ). Differences between means were tested by means of Duncan's multiple-range test [see Kramer (6)]. Differences were significant with the exception of comparisons among Categories 1 through 5, 4 vs. 6, and 5 vs. 6. It should be noted that students who earned degrees did not differ with respect to inner direction and other direction from those who failed to complete their work ( $t = 1.55$ ), though the trend was in the direction of greater inner direction for students who completed degrees.

#### D. DISCUSSION

Before comparing the results of the present study with those reported earlier, it is necessary to comment on the nature of the test instruments used in the assessment of inner direction and other direction. The instrument developed by Bell was composed of items selected on the basis of internal consistency and validated as a total test against a variety of criteria. Kassarian's I-O Scale was also carefully developed with individual items retained on the basis of high internal consistency, and the total test was validated against both external criteria and the differential performance of preselected groups. In each instance items were based on Riesman's descriptive material. Centers used "twenty-five of the most efficient and discriminating items" from Kassarian's scale. Of the 68 items composing Peterson's scale, 47 were adapted from Kassarian's original, unselected items; the remainder, from four other sources. These items were not examined for internal consistency or validity. Kallen measured inner direction and other direction (a) by administering Charles Morris' *Paths of Life* and weighting each choice by the degree of inner direction or other direction involved as estimated by four independent raters and (b) by analyzing the responses to questions concerning the ways in which the subject made decisions. Clearly, the most adequately developed test instruments are those reported by Bell and Kassarian.

##### 1. Sex

The results of the present investigation concerning sex differences in inner directedness other directedness resemble those reported by Kassarian and Centers. Centers found a significant relationship ( $\chi^2 = .05$ ); Kassarian, a slight insignificant trend among both undergraduate and graduate



students; and we found a significant relationship ( $t = 3.67$ ). In each instance men were more inner directed than women. Peterson and Kallen found no significant relationship, though the latter reports a trend toward other direction among males for the measure based on Morris' *Paths of Life*.

## 2. Age

It can be questioned whether a satisfactory test has yet been made of the relationship between age and inner direction, and other direction. In the Kassarian and Peterson investigations, as well as in the present study, no significant relationships were found. However, the samples studied (university students) tended to be relatively homogeneous with respect to age. Centers' study of a cross-section of adults residing in Los Angeles reveals a slight though significant relationship between age and inner direction *vs.* other direction with the younger subjects tending to be more other directed. In no instance has a longitudinal approach been used, with the result that the meaning of any obtained relationship is equivocal.

## 3. Home Town

Kassarian reports no significant relation between inner direction *vs.* other direction and size of home town among undergraduate students and a moderate though significant relationship among graduate students, with the highest relative frequency of inner direction occurring among students from small towns. Peterson found a greater tendency for other-directed subjects to come from large urban centers. In the present study no significant relation is found. The trend toward greater other direction among small-town subjects is inconsistent with that reported by Peterson and Kassarian. No explanation of the conflicting results is available.

## 4. Area of Specialization

There is considerable agreement among the findings of Kassarian, Peterson, and the present study. Service-oriented specializations (ancillary medical specialties, business, and education) tend to be engaged in by other-directed individuals, whereas those specialties that do not involve or stress interpersonal relations (natural science, engineering, and the humanities) tend to attract inner-directed persons. The only important exception occurs in Peterson's study in which inner-directed individuals are overrepresented in social work.

## E. SUMMARY

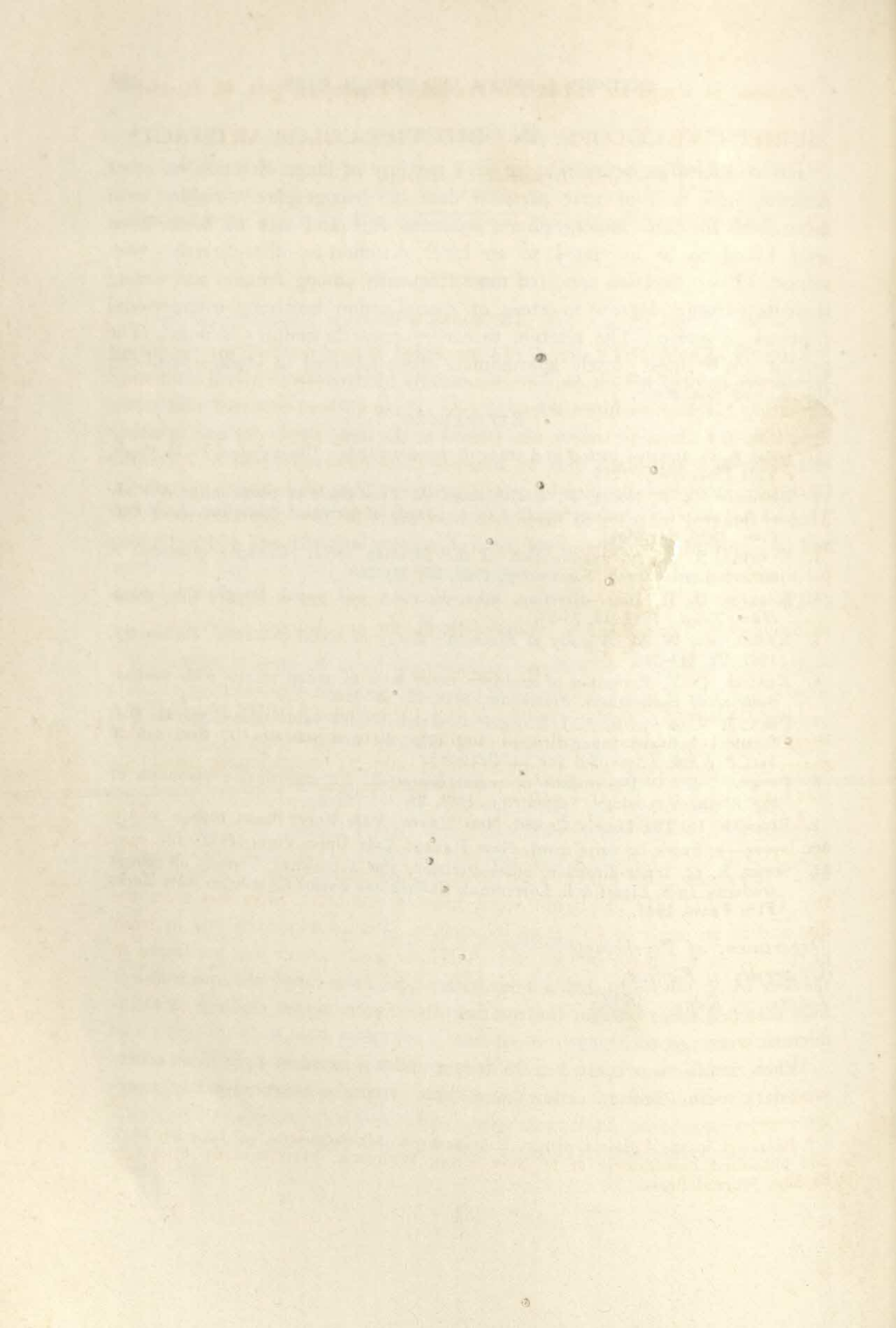
The relationships between score on a measure of inner direction *vs.* other direction and each of four personal data or demographic variables were determined for 226 undergraduate students. Age and size of home town were found to be unrelated to an inner direction-*vs.*-other-direction continuum. Other direction appeared most frequently among females and among students earning degrees in areas of specialization involving interpersonal relations or service. The relation to earlier research findings is noted. The present results most closely approximate those reported in studies involving the Kassarian scale.

## REFERENCES

1. BELL, E. G. Inner-directed and other-directed attitudes. Unpublished Ph.D. thesis, Yale University, 1955.
2. BRODBECK, A. J., NOGEE, P., & DiMASCIO, A. Two kinds of conformity: A study of the Riesman typology applied to standards of parental discipline. *J. of Psychol.*, 1956, **41**, 23-45.
3. CENTERS, R. A. An examination of the Riesman social character typology: A metropolitan survey. *Sociometry*, 1962, **25**, 231-240.
4. KALLEN, D. H. Inner-direction, other-direction and social integration setting. *Hum. Relat.*, 1963, **16**, 75-88.
5. KASSARIAN, W. M. A study of Riesman's theory of social character. *Sociometry*, 1962, **25**, 213-230.
6. KRAMER, C. Y. Extension of multiple range tests to group means with unequal number of replications. *Biometrics*, 1956, **12**, 307-310.
7. PARK, J. N., & SMITH, A. J. Recognition thresholds for value-related words: Differences between inner-directed and other-directed subjects. *J. Personal. & Soc. Psychol.* (Accepted for publication).
8. PETERSON, R. A. Dimensions of social character: An empirical explanation of the Riesman typology. *Sociometry*, 1964, **27**, 194-207.
9. RIESMAN, D. *The Lonely Crowd*. New Haven: Yale Univ. Press, 1950.
10. ———. *Faces in the Crowd*. New Haven: Yale Univ. Press, 1952.
11. SOFER, E. G. Inner-direction, other-direction, and autonomy: A study of college students. In S. Lipset & L. Lowenthal, *Culture and Social Character*. New York: Free Press, 1961.

Department of Psychology  
University of Kansas  
Lawrence, Kansas 66045





## SUBJECTIVE COLORS: AN OBJECTIVE-COLOR ARTIFACT\*

*Psychology Department, Dalhousie University*

---

B. J. FROST

---

### A. INTRODUCTION

Recently, Knehr and Lorenz (1) presented a new method for producing "subjective colors," which consisted essentially of stroboscopically illuminating a rotating black-and-white patterned disc. These authors reported that when light from a fluorescent source was pulsed at the same frequency as a rotating half-white and half-black disc, an orange-yellow color appeared to cover the black section of the disc, while a well-saturated blue was seen to precede it. Monochromator matches of these two hues made by two observers were 590 and 601 millimicrons for the yellow-orange color and 481 and 491 millimicrons for the blue.

The study to be described here presents evidence that indicates that some objective color component is produced by Knehr and Lorenz's method.

### B. PROCEDURE

The experiment consisted essentially of a replication of Knehr and Lorenz's study; but (in addition) color photographs were taken of the disc under observation conditions, and a comparison of subjects' spectrometer matchings was made with the spectral output of the light source.

A 12-inch diameter black cardboard disc with a seven-inch side white equilateral triangle in the center was rotated at approximately 3600 rpm by a Gerbrands color mixer. Two 40-watt fluorescent tubes (General Electric Type F20T12D Daylight) were oriented vertically, three feet from the disc and at an angle of 45° from its surface. Ss were seated directly in front of the disc at a distance of two feet. A Hilger spectrometer was positioned to the left of S, who adjusted a micrometer-type drive until the spectrometer hues matched those seen on the rotating disc. Spectrometer readings in millimicrons were recorded.

When the disc was rotated at 3600 rpm under fluorescent light in an otherwise dark room, "brownish-yellow" and "blue" triangles appeared partly super-

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on June 29, 1965, and published immediately at 35 New Street, Worcester, Massachusetts Copyright by The Journal Press.



imposed on each other to form a star shape, as illustrated in Figure 1. The brownish-yellow triangle (abc) appeared more prominent than the blue triangle (def), while the area between the points and cusps of the star appeared achromatic. Under both tungsten and fluorescent illumination Annulus A appeared black with a dark Mach band at the inner edge, and Disc B appeared white with a light Mach band around its circumference; however, under tungsten illumination only an achromatic radial gradient from black to white appeared across the inner annulus.

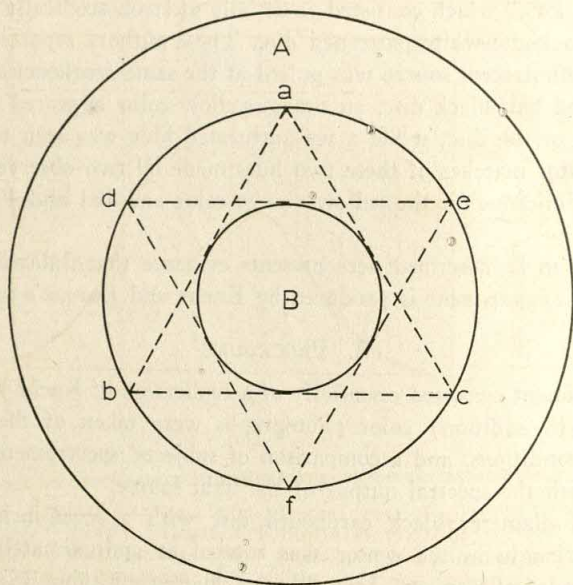


FIGURE 1

SCHEMATIC DRAWING OF DISC AS IT APPEARED WHEN ROTATING AT 3600 RPM. UNDER FLUORESCENT ILLUMINATION.  
(Triangle abc appeared a brownish-yellow color, while triangle def appeared blue.)

Five male graduate students served as subjects. Each was instructed to disregard as far as possible differences in saturation and brightness between the disc and spectrometer hues, while adjusting the spectrometer until it appeared the same as the disc hue *E* had indicated *S* should match. Color names were not used (in order to avoid the possibility of biasing *S*'s settings). Two practice settings followed by 10 experimental settings were performed by each subject for both the "blue" and "yellow" hues. The spectrometer starting position was randomly set at either 420 or 680 millimicrons.

TABLE 1  
MEAN SPECTROMETER SETTINGS (IN MILLIMICRONS) FOR THE  
BLUE AND YELLOW DISC HUES

Subject	"Blue"	Hue "Yellow"
A	486	610
B	469	605
C	485	605
D	472	612
E	478	601
Mean	478	606.6

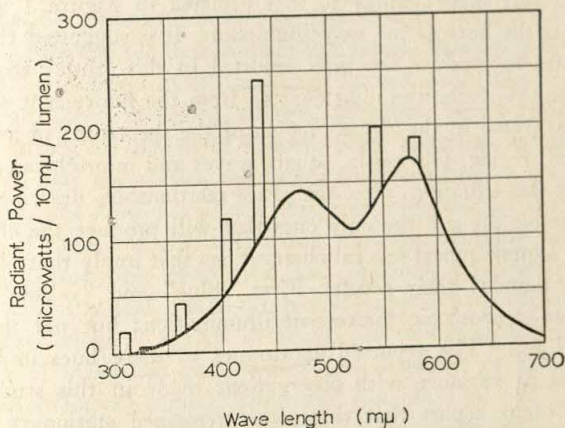


FIGURE 2

RELATIVE-ENERGY OUTPUT OF A GENERAL ELECTRIC  
40-WATT DAYLIGHT FLUORESCENT TUBE.  
(The smooth curve represents light resulting from phosphor  
excitation; the bars, visible light generated directly  
by the mercury arc—courtesy Canadian  
General Electric Company.)

### C. RESULTS

Table 1 shows Ss' mean spectrometer settings in millimicrons for both the blue and yellow disc hues. The relative energy output of the fluorescent light source used to illuminate the rotating disc is presented in Figure 2.

### D. DISCUSSION

The most compelling evidence for the objectivity of the hues matched in this experiment is provided by their appearance in color transparencies of the rotating disc. Unfortunately, these photographs cannot be reproduced; but the hues observed in the photographs were very similar to those that appeared in the disc.



When the relative energy output of the fluorescent source used to illuminate the disc is compared with Ss' spectrometer matches of the two hues seen in the disc, fairly close agreement is found; however, Ss' matches of the "yellow" section of the disc are slightly displaced toward longer wavelengths.

The segregation of these two hues into different regions of the rotating disc can be accounted for as follows. If square waves of white light are pulsed at 60 cps onto a disc such as the one used in this experiment, a white triangle appears stationary at 3600 rpm: the normal stroboscopic effect. If two square waves of different monochromatic light half a cycle out of phase replace the white light, a star shape similar to that depicted in Figure 1 appears with adjacent points the hues of the two illuminants. It is suggested that a similar process operates to produce the hues reported in this study: that is, during one stage of an AC cycle, the light emitted from the fluorescent source is predominantly saturated in the "blue" region of the spectrum; in another stage, in the "yellow" region. Obviously, square waves and monochromatic light are not present in this situation; also, the phase relationships described in the explanatory example are not the only ones that will produce the effect.

Knehr and Lorenz report several observations that imply that objective hues may have confounded their results. First, "blue" and "yellow" hues were obtained under stroboscopic fluorescent illumination; but not under stroboscopic white light. (The reported appearance of these hues under tungsten illumination is at variance with observations made in this study.) Second, Knehr and Lorenz report that these hues remained stationary when light pulses were synchronized with disc rotation, thus showing the relevance of the stroboscopic effect. Third, monochromator matches by their two observers closely resemble data gathered in this experiment in which an objective color component has been found.

The results of this study indicate that subjective-color experiments may be confounded by the presence of objective colors when fluorescent light sources powered by AC current are used; consequently, the author suggests that a white-light source of known relative-energy output, powered by DC current, be used in future experiments.

#### REFERENCES

1. KNEHR, C. A., & LORENZ, R. J. Subjective color: A new method of producing the phenomenon. *J. of Psychol.*, 1964, **58**, 353-356.

*Psychology Department*  
*Dalhousie University*  
*Halifax, N. S., Canada*

## RECENT ADVANCES IN THE PSYCHOLOGY OF MASCULINITY AND FEMININITY\*<sup>1</sup>

*Department of Psychology, Xavier University*

---

VYTAUTAS J. BIELIAUSKAS

---

### A. INTRODUCTION

While the fact that mankind consists of two sexes, male and female, is obvious and has been accepted since time immemorial, the question concerning the differences between the human male and female also has a history extending to the earliest known existence of the human race. The question about sexual differences has intrigued writers, scientists, poets, theologians, and philosophers during all phases of history. Some parts of this question have been related to new discoveries and better understanding of human anatomy and physiology, and they have been answered progressively better with the advancement of these branches of science. However, it seems that while the knowledge concerning the anatomical and physiological differences of sexes may interest curiosity and desire for factual information, the questions concerning the psychological aspects of sexual differences appear to include a certain aspect of mysticism and mystery. As a matter of fact the question concerning psychological sexual differences has been, is, and probably will be asked by everyone at different ages and very likely without reaching a satisfactory answer.

The problem of psychological differences between men and women was attacked in using scientific approach for the first time by Havelock Ellis (8). His research and his suggestions have been evaluated, criticized, and questioned by many authors since then without reaching a more definite conclusion [*cf.*(5)]. While there always has been agreement among researchers that there is a difference between the thinking, feeling, and acting of men and women, the reasons for these differences have been questioned without reaching conclusions. In a paper presented to the Catholic International Congress of Medical Psychology (5), the author reviewed many of the investiga-

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on July 1, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was supported by the National Science Foundation Grant GU-408. The paper was presented as an invited lecture to Barat College of the Sacred Heart, Lake Forest, Illinois, on February 3, 1965.



tions concerning the psychological characteristics of maleness and femaleness and came to the conclusion that there was little agreement on the meaning of the concepts of masculinity and femininity. Suggestion was made that the meaning of sexual differences tends to depend upon cultural interpretation and the particular phase of history in which one lives. However, in spite of the lack of overall agreement on specific definitions of masculinity and femininity, most writers seem to assume that psychological sexual differences do exist.

Even within recent decades, numerous research publications have suggested that investigators still find a need for clarification and further empirical evaluation of concepts dealing with sex differences. Indeed, various aspects of these differences have been subjected to empirical evaluation; and some questions that earlier were answered on the basis of speculation and personal projections can now be attacked on the basis of objective data. However, since most of the empirical research findings dealing with various aspects of masculinity and femininity are fragmentary rather than global, it is difficult to benefit adequately from them in attempting to develop a holistic point of view in regard to masculinity and femininity. Therefore, there exists a need to survey the most recent advances in psychology of masculinity and femininity to establish their interrelations and to examine them in terms of earlier theories. To meet this need is the goal of this paper.

#### B. DEVELOPMENT OF SELF-CONCEPT AND IDENTIFICATION

One of the problems that receives frequent attention from investigators dealing with sexual differences is the question concerning the development of the self-concept ("who am I"), and a special feature of the self-concept is the question concerning sexual identification. Both the self-concept and sexual identification the child learns early from his environment. To what is learned the child responds with certain behavior, and the observation of such behavior gives information concerning the incorporation of learned materials into personality structure.

Children very early learn their roles as boys and girls. These roles are assigned just after birth and are continuously reinforced as the children grow older; thus the overt behavior of children should indicate that beginning at a certain age they should be able to distinguish between male and female roles and demonstrate this knowledge through action. Hartley and Hardesty (10) have reported that children 8 to 11 years of age perceive clear-cut differences in sex roles. Sex-role activities in childhood seemed to parallel specific traditional female-male activities at the adult level: for example,

plays with doll carriage, does dishes, dusts table, helps mother to hang clothes, etc., for girls; and plays with toy air rifle, carries wood into the house, helps man fixing ceiling, etc., for boys. Similar though not identical results were reported by Bieliauskas (4), who studied developmental patterns in sexual identification via children's drawings of the human figure. In that study, the House-Tree-Person (H-T-P) drawings produced by 10,650 children (5 to 15 years of age) indicated that as a rule each sex prefers to draw a person of his own sex; and this tendency seems to increase with age, particularly after the age of 9. While girls show a clear preference for the female figure at the age of 6, boys reach a clear-cut preference only at the age of 9 and thereafter. While the boys remain stable in their preference to draw the male figure at the age of 9 and above, girls show a pronounced reduction in their preference to draw the female figure at the age of 12. By implication, the foregoing can be interpreted to mean that some confusion in sexual identification takes place in some girls at the age of 12 (which is the age at which girls begin to experience themselves more consciously in their female role than theretofore). Brown (6), referring to similar findings, suggested that such results provide support for the theory that sexual identification is influenced by culture. In Western culture, which values the male sex over the female sex, the female tends to show more difficulty in defining her sexual identification—especially at the phase in development when she becomes aware of belonging to the culturally nonpreferred sex. This experience, however, according to Bieliauskas (4), disappears later on and seems to have little measurable bearing upon later development.

While boys show more stability than girls in sexual identification, boys also experience problems that seem to be sex-linked, as indicated in a study by Rosenberg and Sutton-Smith (18). These investigators examined the effects of ordinal position on sex-role identification and anxiety. They studied children in one-child, two-child, and three-child families—ages 9-12—and found that single male children were highly anxious and that single female children were low on anxiety. Also, "in three-child families, the male child with two female siblings shows heightened masculinity and anxiety (conflict induced)." The same was not true for girls. In general, the results imply that "children in one-, two-, and three-child families are differentially affected by the reinforcements in sex-role learning afforded by parents and siblings."<sup>2</sup>

An interesting observation in reverse is supplied in a recent study by Lansky

---

<sup>2</sup> As Rosenberg and Sutton-Smith limited the family tree to a maximum of three, the question remains unanswered as to what happens with the anxieties of the male children in larger families.



(13), who found that sex-role identification (as measured by masculinity-femininity tests) tends to depend upon the predominance of male or female children in the family. It seems that male-female identification is not something that children learn from parents; but that parents, too, are affected in their identification by their children. In the latter sense, psychological-sexual identification is a continuous process.

### C. INTEREST, OCCUPATION, ACHIEVEMENT, AND PERFORMANCE

Terman and Miles (19), in one of the first extensive studies of masculine and feminine traits in American culture, reached conclusion as follows:

From whatever angle we have examined them the males included in the standardization groups evidenced a distinctive interest in exploit and adventure, in outdoor and physically strenuous occupations, in machinery and tools, in science, physical phenomena, and inventions; and, from rather occasional evidence, in business and commerce. On the other hand, the females of our groups have evidenced a distinctive interest in domestic affairs and in aesthetic objects and occupations; they have distinctly preferred more sedentary and indoor occupations, and occupations more directly ministrative, particularly to the young, the helpless, the distressed. Supporting and supplementing these are the more subjective differences—those in emotional disposition and direction. The males directly or indirectly manifest the greater self-assertion and aggressiveness; they express more hardihood and fearlessness, and more roughness of manners, language, and sentiments. The females express themselves as more compassionate and sympathetic, more timid, more fastidious, and aesthetically sensitive, more emotional in general (or at least more expressive of the four emotions considered), severer moralists, yet admit in themselves weaknesses in emotional control and (less noticeably) in physique (19, pp. 447-448).

The foregoing statement concerning masculine and feminine characteristics is probably one of the few available based upon empirical study. Although Terman and Miles' study has been criticized by various authors [*cf.* (9, 17)], some of the observations made require comment and evaluation in terms of contemporary research.

There have been various attempts to establish interest profiles typical of women; others, typical of men. And all attempts are based on the assumption that interests and occupational choices are sex linked. As examples, one can mention Strong's Vocational Interest Blank (which has two special forms: one for men, one for women) and provides two different occupational-interest profiles. Recently, Overall (15) has suggested a masculinity-femininity scale for Kuder's Preference Record. Research on both scales implies that individuals taking these tests consciously or unconsciously attend to activities that

discriminate between males and females. However, times change; so do interests.

Rosenberg and Sutton-Smith (17) have pointed out that some changes have occurred in interests in children's games since Terman and Miles' study, and that the changes reflect certain sex-linked behavior. As for specific interests and occupational choices among adults, Terman and Miles findings are still in effect. Johnson and Knapp (11), after having evaluated sex differences in aesthetic preferences, concluded that "men and women do differ significantly in their preferences for art of all types"; however, the sex differences were less significant than "differences in preference caused by such variations among groups as age, social class, special training and vocation." A statement from the foregoing study—despite the fact that it goes beyond the data—is interesting: "It may be suggested that in terms of *content* women tend to prefer the personal rather than the impersonal, the intimate rather than grand, the sensuous rather than abstract." This statement almost parallels the one cited from Terman and Miles' study and can even be compared with certain suggestions advanced by Ellis (8).

Various investigators have noted sex-linked differences for performance and achievement. As a rule, the earlier observations concerning *better* female performance find support in recent research publications. Ames and Ilg (4) show evidence that "test performance of girls in the five- to nine-year-old range appears to be considerably advanced over that of boys in the same age range." Further, they suggest that "boys might benefit from entering school at slightly later age than girls." On the other hand, some culturally expected patterns in motivation and interest in regard to specific subjects still exist: For example, boys show higher motivation and better achievement than girls in mathematics (16); learning and behavior disorders are three to 10 times more frequent among males than females of the same chronological age (3); and Cattell and Lawson's findings (7) imply that there is evidence of sex-linked characteristics in group performance on aspiration level and persuasability.

Weiss (20), who studied the behavior of women in so-called masculine activities, found that college women tend to limit their achievement when competing with college men; thus avoid the disruption of the rules governing heterosexual encounter.

#### D. CULTURAL INFLUENCES

The foregoing studies almost unanimously imply that masculinity and femininity are characteristics greatly influenced by culture, and additional support for this assumption is provided by certain cross-cultural studies.



Klausner (12) studied the sexual behavior of new immigrants in Israel and showed that their sexual behavior tended to approach quickly that of the receiving population. In another study, Arkoff, Meredith, and Iwahara (2) found significant differences between Japanese and Japanese-American groups in terms of the behavior of males and females. While cultural influences upon masculine and feminine behavior are obvious, one cannot overlook the fact that even after taking cultural differences into account, one still finds differences in masculinity and femininity in each of several cultures. As Bieliauskas (5) has pointed out, "... we may investigate and argue about the causes for sexual personality differences, but we cannot deny that they exist."

#### E. MASCULINITY-FEMININITY SCALES

There are two basic difficulties in using the empirical approach to the study of masculinity and femininity: (a) the difficulty in defining masculinity and femininity and (b) the problem of developing instruments to assess the characteristics. The difficulties in definition have been illustrated in the descriptions of various studies on the subject. As far as the measuring instruments of masculinity and femininity are concerned, we have relied primarily upon various psychological M-F (masculinity-femininity) scales, most of which depend upon verbal description of interests or activities that Western culture considers as male, female, or neutral. For example, one M-F scale requires that S mark "true" or "false" statements such as: "I think I would like to be a librarian," "I enjoy reading love stories," and "I like mechanics magazines."

There is no doubt that certain of activities can be considered feminine and others masculine; but in many cases the descriptions are so obvious that one would have no difficulty in answering the questions the way one's cultural stereotype implied they be answered; so give little insight into the masculine-feminine aspects of S. Some scales deal primarily with vocational-occupational interests; but they, too, are obviously culturally dependent. In 1949, Franck and Rosen (9) devised a masculinity-femininity test that avoided verbal material and consisted primarily of drawing completions that (according to various investigations) are accomplished in different manners by men and women. The validity and reliability of this test has been highly endorsed in a recent publication by Lansky (13). It seems that this test has some promise, although difficulties are encountered in scoring and evaluation.

Nichols (13) developed a scale that consists of items showing sex differences of which people generally are not conscious; but the scale was found to have low internal consistency, and its applicability is questionable. There

is no doubt that there are changes in times and changes in cultural interpretations of masculinity and femininity. Therefore, various M-F scales based upon certain assumptions valid 25 years ago may need revision in light of present-day social and cultural changes.

#### F. NEED FOR NEW APPROACHES AND NEW PHILOSOPHY OF MASCULINITY AND FEMININITY

As the research findings indicate, there have been changes in the basic thinking concerning masculine and feminine characteristics in Western culture. Certain occupations and certain activities that could not be chosen by women earlier can be chosen now. Activities that may be considered masculine in one culture may be feminine in another. The fact that more women than ever before now occupy professional and semiprofessional positions cannot be overlooked as an important factor that eventually will produce changes in cultural stereotypes of sexual differences.

Probably one of the reconsiderations that will be needed is one of viewing masculinity and femininity not as two competitive characteristics but as complementary personality traits. Psychological characteristics usually tend to overlap rather than be exclusive; therefore, it is possible that masculinity and femininity can overlap in the same person without producing personality distortion. Piotrowski, in a personal communication (1961), has indicated that men who are successful therapists show more feminine characteristics than expected on M-F scales. In evaluating his findings, Piotrowski discovered that one characteristic that increased the feminine compound in M-F score was that what popular opinion would call "intuition": i.e., the ability to observe subtle behavior quickly and interpret it accurately. This ability, one of a clinician's chief assets, a "feminine" characteristic.

Does a woman engaged in a so-called "male" occupation have to become masculine or can she remain feminine in spite of her occupation? Some personal observations, which are also supported by the results of some pilot studies, imply that some women engaged in so-called "male" occupations are aware of a need to remain feminine and (on M-F scales) show more feminine traits than do women working in typical "female" occupations.

#### G. SUMMARY

The author has reviewed some recent studies dealing with various aspects of masculinity and femininity. Masculinity and femininity, as psychological experiences, are developed in early childhood and are related to the development of the self-concept and to the achievement of identification. Sexual dif-



ferences *do* exist in terms of sexual identification. Women and men show differences in interest, occupational choice, achievement, and performance; but most of the differences are influenced by the culture in which one lives. But, even after taking cultural differences into account, one still finds differences in masculine and feminine behavior. To evaluate psychological-sexual differences, various M-F scales have been devised, but the validity of these has been questioned. As changes in cultural emphases require changes in stereotypes of masculinity and femininity, the author suggested a complimentary approach rather than a competitive approach in future studies of masculinity and femininity.

While question has been raised concerning the value of stereotypes sexual differences, there cannot be any question as to the fact that sex differences exist. In reality, every human being exists not as a human being but as a man or as a woman. For a man to be a man, and for a woman to be a woman is to be what one is supposed to be. Philosophically speaking, it is good for a being to be what it is supposed to be. Therefore, for a *man* to be *masculine* and for a *woman* to be *feminine* is that which makes each an authentic human being, and enhances the maturity and the richness of personality.

#### REFERENCES

1. AMES, L. B., & ILG, F. L. Sex differences in best performance of matched girl-boy pairs in the five to nine year old age range. *J. Genet. Psychol.*, 1964, **104**, 25-34.
2. ARKOFF, A., MEREDITH, G., & IWAHARA, S. Male-dominant and equalitarian attitudes in Japanese, Japanese-American and Caucasian-American Students. *J. Soc. Psychol.*, 1964, **64**, 225-229.
3. BENTZEN, F. Sex ratios in learning and behavior disorders. *Amer. J. Orthopsychiat.*, 1963, **33**, 92-98.
4. BIELIAUSKAS, V. J. Sexual identification in children's drawings of human figure. *J. Clin. Psychol.*, 1960, **16**, 42-44.
5. ———. Psychological aspects of femininity and masculinity. Paper given to Catholic International Congress of Medical Psychology, Toulouse, France, July, 1963.
6. BROWN, D. G. Sex-role development in a changing culture. *Psychol. Bull.*, 1958, **4**, 232-242.
7. CATTELL, R. B., & LAWSON, E. D. Sex differences in small group performance. *J. Soc. Psychol.*, 1962, **58**, 141-145.
8. ELLIS, H. *Man and Woman: A Study of Human Secondary Sex Characters*. New York: Scribner's, 1904.
9. FRANCK, K., & ROSEN, E. A projective test of masculinity-femininity. *J. Consult. Psychol.*, 1949, **13**, 247-256.
10. HARTLEY, R. E., & HARDESTY, F. P. Children's perception of sex roles in childhood. *J. Gen. Psychol.*, 1964, **105**, 43-51.
11. JOHNSON, O., & KNAPP, R. H. Sex differences in aesthetic preferences. *J. Soc. Psychol.*, 1963, **61**, 279-301.

12. KLAUSNER, S. Z. Inferential visibility and sex norms in the middle east. *J. Soc. Psychol.*, 1964, **63**, 1-29.
13. LANSKY, L. M. The family structure also affects the model: Sex-role identification in parents of preschool children. *Merrill-Palmer Quart. Beh. Devel.*, 1964, **10**, 39-50.
14. NICHOLS, R. C. Subtle, obvious and stereotype measures of masculinity-femininity. *Educ. & Psychol. Meas.*, 1962, **22**, 449-461.
15. OVERALL, J. E. A masculinity-femininity scale for the Kuder Preference Record. *J. Gen. Psychol.*, 1963, **69**, 209-216.
16. POFFENBERGER, T., & NORTON, D. Sex differences in achievement motive in mathematics as related to cultural change. *J. Genet. Psychol.*, 1963, **103**, 341-350.
17. ROSENBERG, B. G., & SUTTON-SMITH, B. The measurement of masculinity and femininity in children. *Child Devel.*, 1959, **30**, 373-380.
18. ROSENBERG, B. G., & SUTTON-SMITH, B. Ordinal position and sex-role identification. *Gen. Psychol. Monog.*, 1964, **70**, 297-328.
19. TERMAN, L. M., & MILES, C. C. Sex and Personality: Studies in Masculinity and Femininity. New York: McGraw-Hill, 1936.
20. WEISS, P. Some aspects of femininity. *Diss. Abstr.*, 1962, **23**, 1083.

Department of Psychology  
Xavier University  
Victory Parkway  
Cincinnati, Ohio 45207





## SPEECH AVOIDANCE AMONG CHILDREN WITH ORAL-COMMUNICATION DEFECTS\*

*Department of Speech, Northern Illinois University*

---

LOUIS LEREA AND BRUCE WARD

---

### A. INTRODUCTION

A reluctance to speak or an avoidance of situations that require oral participation is not only typical of those who stutter, but is often evident among individuals with other types of communication problems. Johnson (1) has stated that "Even, under the best of circumstances, the child with defective speech is almost certain to experience frustration and anxiety in speaking situations." The clinical significance of the speech-defective child's emotional adjustment has been underscored by Powers (2) who wrote:

This is one of the most essential parts of the clinical study of a case. The person's emotional characteristics and modes of adjustment frequently have a causal relationship to his speech problem and, in any case, are important to the therapeutic management of the problem (2, p. 771).

It is generally assumed that the severity of a child's speech handicap is associated with social rejection and subsequent speech-avoidant behavior. In this study, the authors constructed an inventory of speech avoidance in an effort to determine whether a relationship exists between speech misarticulations and speech anxiety among elementary-school children.

### B. CONSTRUCTION OF THE SPEECH-AVOIDANCE TEST

Preliminary efforts to develop a test that would enable one to appraise a child's reactions to his oral-communication difficulties and the degree to which he expresses his willingness to interact in speaking situations were initiated 10 years ago. Items were prepared that were assumed to be within the repertoire of experiences of most elementary-school children, regardless of sex or geographical origin. Questions were then revised or discarded, and new items were included in the inventory on the basis of responses given by children representing a diverse clinical population. These speech-handicapped or hearing-handicapped children were referred to speech clinicians for di-

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on July 2, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



agnosis or therapy. A final version of the test, consisting of 49 items, was completed in 1963. Depending upon the sex of the examinee, each item is followed by the question, "Are you like [name of boy or girl]?" This format was selected because it seemed concrete and credible to the child. The test was prepared so that the 49 items could be (a) given to groups as well as individuals, (b) scored quickly, and (c) interpreted easily. The instructions for the test were as follows:

We will be reading about boys [girls]. Some of the boys [girls] feel certain ways about things. For example, John [Mary] likes to play. After we read about a boy [girl], you will be asked: Are you like John [Mary]? If you are like the boy [girl] in the story, you should put a circle around the word YES. If you are not like the boy [girl] in the story, you should draw a circle around the word NO. There are no right or wrong answers because you know if you are like the boys [girls] in the stories or if you are not like the boys [girls] in the stories.

An item analysis of the test was completed from the responses of 101 children. Fifty-five were diagnosed as having severe articulation difficulties (nine with cleft palate, one with cerebral palsy, nine with hearing loss, and 36 with multiple functional defects); and 46 were judged to have mild functional articulatory errors. The organically handicapped children were under the care and sponsorship of the University of Illinois, Division of Services for Crippled Children, and were seen by diagnostic teams consisting of pediatricians, neurologists, orthodontists, psychologists, audiologists, and speech therapists. In this study all children classified as functional articulatory defectives possessed speech problems of undetermined origin and in all other respects did not seem to depart significantly from the norm for their age group.

A picture-articulation test representing the major phonemes of general American speech and a confirming test of connected speech were employed in evaluating the severity of the child's articulation defects. The children designated as severe misarticulated four or more consonants that initiated and closed a syllable or a minimum of three phonemes, two of which were *s*, *r*, or *l* sounds. The group considered as mild demonstrated articulation errors in their production of not more than two consonants. The mean age of the 37 boys and 18 girls comprising the severe group was 8:2; the group with mild speech defects consisted of 20 boys and 26 girls, and their mean age was 8:3. The age range of these 101 children was 6 to 13 years.

The responses of the mild group and severe group were tabulated in terms of the number who answered yes or no to each of the 49 items. A contingency table was prepared, and the chi-square statistic (with Yates' correction for

continuity) was computed to determine the discriminative value of each item. Those questions that differentiated between the groups beyond the .05 level of confidence were incorporated in the revised form of the Speech-Avoidance Test. Five of the 30 items that met the foregoing criterion are as follows:

Joe [June] points to the candy he [she] wants to buy instead of telling the storekeeper what he [she] wants. Are you like Joe [June]?

Peter [Alice] likes to play alone because children cannot understand him [her] when he [she] talks to them. Are you like Peter [Alice]?

George [Jill] likes to tell his [her] family about things that happen at school. Are you like George [Jill]?

When someone asks Marvin [Mary] a question, he [she] waits for his [her] parents to answer for him [her]. Are you like Marvin [Mary]?

Mike [Helen] sometimes crosses the street so he [she] will not have to talk to children he [she] knows. Are you like Mike [Helen]?

### C. RELIABILITY

The reliability of the Speech-Avoidance Test was evaluated by the test-retest method. It was given to 56 boys and 48 girls with mild or severe functional articulation defects. The mean age of each group was 7:9 (range-6:4 to 11:1). A period of two weeks to one month elapsed between the first and second administration. Person product-moment coefficients of correlation were .77 for boys, .84 for girls, and .80 for all cases.

### D. VALIDITY\*

The authors hypothesized that children with severe communication handicaps would respond differently to the 30-item inventory than would those with mild speech problems. Two separate validation studies were undertaken.

First, the completed test was given to 36 children (18 boys and 18 girls) with mild functional defects and to 36 children (18 boys and 18 girls) with severe functional articulation defects. The mean ages of the children classified as severe and mild were 8:2 and 8:3 respectively.

A two-by-two factorial design was employed to determine whether or not the 30-item test differentiated between the groups and the sexes. The *F* ratio for groups is significant beyond the .01 level [see Table 1]. Children di-

TABLE 1  
FIRST ANALYSIS OF VARIANCE

Source	<i>df</i>	<i>MS</i>	<i>F</i>
Groups	1	709.38	42.38*
Sex	1	13.00	.78
Interaction	1	28.90	1.73
Error	68	16.74	

\*  $p < .01$ .



agnosed as severe functional articulatory defectives expressed a greater number of emotive and avoidant responses than did children whose misarticulations were limited to one or two minor consonant errors.

Second, three groups of elementary-school children were recruited as subjects. Groups 1 and 2 were children with mild or severe articulation defects considered to be functional in character. Group 3 consisted of 20 hard-of-hearing children with associated severe articulation problems. A Beltone 15C audiometer was utilized to determine the extent of hearing loss among these children. The three-frequency technique for averaging hearing loss in the better ear for each of these children revealed that their mean hearing loss was 47-dB (1951, ASA Standard).

The mean ages of the boys and girls with mild and severe functional speech defects were both 9:1; the average age of the organically handicapped children was 10:0 years.

The test was administered to the foregoing children individually or in groups; the means and standard deviations are shown in Table 2. The girls consistently manifested greater variability than the boys.

TABLE 2  
MEANS AND STANDARD DEVIATIONS FOR THE THREE GROUPS AND SEXES

Parameter	Mild defects		Severe defects		Hard-of-hearing children	
	Boys ( <i>N</i> = 10)	Girls ( <i>N</i> = 10)	Boys ( <i>N</i> = 10)	Girls ( <i>N</i> = 10)	Boys ( <i>N</i> = 10)	Girls ( <i>N</i> = 10)
Mean	3.50	3.00	9.80	9.40	11.50	12.70
<i>SD</i>	2.55	5.33	3.12	6.17	4.77	7.30

The results of a three-by-two analysis of variance are shown in Table 3. The *F* values for sex and the interaction effect (as in Table 1) are inconsequential. The *F* ratio for groups, however, is 15.26 and is significant beyond the .01 level. These hard-of-hearing children, perhaps because of the marked

TABLE 3  
SECOND ANALYSIS OF VARIANCE

Source	<i>df</i>	<i>MS</i>	<i>F</i>
Group	2	416.32	15.26*
Sex	1	.15	.01
Interaction	2	4.55	.17
Error	54	27.28	

\*  $p < .01$ .

severity of their communication problems, express a greater number of adverse and avoidant responses than do the children with severe functional articulation handicaps. Again, sex seems not to be a critical determinant of speech avoidance among these elementary-school groups.

### E. DISCUSSION

The therapist is acutely aware of the need for information concerning the child's internalized reactions to his oral-communication disorder. The premise is that evasion from speaking situations is associated with the severity of the child's speech problem. The authors constructed a short, flexible, and simple test of speech avoidance that can be given individually or in groups. Testing time does not exceed 25 minutes.

The test responses of children with minor speech defects differed significantly from those of the more-seriously defective children. In this population, therefore, the severity of a speech handicap seems to be related to speech avoidance. This result suggests that anxiety, manifested by expressed emotive and avoidant responses in speaking situations, may be operant among the children with severe functional articulation defects and hearing deficits. External criteria for evaluating the significance of an *individual's* speech-avoidance score must be established. Furthermore, the application of this test to other populations of children with oral-communication handicaps should be determined and norms compiled.

Therapists are in agreement that the relationships between parents and child are significant parameters to be considered in diagnosis and therapy. The Speech-Avoidance Test could be administered to the parents by simply rewording the question following each of the 30 items so that it reads, "Is your child like [name of boy or girl]?" Multiple comparisons between the responses of the mother, father, and child may permit the therapist to abstract valuable clinical information.

### F. SUMMARY

This investigation was concerned with the relationship between the severity of speech misarticulations and speech avoidance among elementary-school children. A test was constructed to assess speech avoidance. An item analysis of the responses of 101 children with mild or severe articulatory defects yielded 30 items indicative of speech avoidance. The test-retest reliability is .80, and the results of the two validity studies show that the Speech-Avoidance Test effectively differentiates between criterion groups.



## REFERENCES

1. JOHNSON, W. Speech Handicapped School Children. New York: Harper, 1956.
2. POWERS, M. H. Clinical and educational procedures in functional disorders of articulation. In L. Travis (Ed.), *Handbook of Speech Pathology* (Chap. 24). New York: Appleton, 1957.

*Department of Speech*  
*Northern Illinois University*  
*DeKalb, Illinois*

## THE EFFECT OF PREVIOUSLY LEARNED VERBAL HABITS UPON TWO RESPONSE CLASSES IN VERBAL CONDITIONING\*<sup>1</sup>

*Department of Psychology, University of Kentucky*

PAULINE R. CLANCE AND T. R. DIXON

### A. INTRODUCTION AND PROBLEM

The primary emphasis of considerable research in verbal conditioning has been directed toward the relationship between various reinforcing stimuli and verbal-response change. Until recently, little attention has been focused upon the effect of previously learned verbal habits in such conditioning. In 1965, Dixon reported that the verb used in a sentence affects conditioning of the personal pronouns "I" and "we" (2). In a later study (3) Dixon found that the verb-pronoun connection was not affected by the E-S relationship, while reinforcement was. In both experiments, Dixon used a modification of a sentence-completion procedure in which the verb, which was presented with six pronouns, was an independent variable and the pronouns "I" and "we" were dependent variables. These investigations differ from studies, such as those of Buss and Durkee (1), Simkins (3), and Ferguson and Buss (5), in which the verb was a dependent variable; i.e., Dixon's primary emphasis was upon the effect of previously learned verbal habits in verbal conditioning, while the primary emphasis for Buss and others was upon the relation of reinforcement and hostility.

Dixon's results indicate that the personal pronouns "I" and "we" are used more frequently with good-impression verbs than with bad-impression verbs and show the verb-pronoun relationship to be a type of verbal habit, with the verb functioning as a discriminative stimulus for the pronoun response. Further, the data indicate that such verbal habits are established prior to experimental reinforcement (reflected in the operant period) and continue to affect the pronoun-response frequency in the experimental-acquisition period.

Because Dixon's experiments involve the relation between verbs and a relatively limited class of responses (i.e., "I" and "we"), the present experi-

\* Received in the Editorial Office, Provincetown, Massachusetts, on July 7, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This research was supported in part by Public Health Service Research Grant MH08884-01.



ment was designed to determine if the verb effect can be extended to other responses: specifically, to the impersonal pronouns "he" and "they." The authors reasoned that the impersonal pronouns "he" and "they" constitute a response class very different from that comprising the personal pronouns "I" and "we." "I" and "we" are self-referents and *S* is, by context, referring to himself when using those pronouns; "he" and "they" are other referents in that *S* refers to others when using them. If the verb systematically affects only *S*'s choice of self-referent pronouns, then the significance of such pre-experimentally acquired verbal habits for sentence-completion verbal conditioning would be limited to indicating that the verb is a discriminative stimulus when *S*s refer to themselves; however, if the verb affects *both* self-referent pronouns and other-referent pronouns in the operant period and in the acquisition period, then one can conclude that the verb in a sentence is an effective and well-established discriminative stimulus for a variety of subject pronouns. If this proves to be the case, one should include the effects of all such previously learned verbal habits in any complete account of this type of verbal conditioning.

## B. METHOD

Eighty female volunteers from registered nursing-degree programs served as subjects. Four experimental conditions were used: (a) "I" and "we" reinforced, (b) "he" and "they" reinforced, (c) "I" and "we" control, and (d) "he" and "they" control ( $N = 20$  per condition). The operant period consisted of two blocks of 30 trials each, during which no reinforcement was given. The acquisition period consisted of two blocks of 30 trials each, in which *E* reinforced the use of the personal pronouns "I" and "we" in Condition (a) and the use of the impersonal pronouns "he" and "they" in Condition (b), by saying "good" on a 100 per cent schedule.

The procedure was similar to that reported by Dixon (2). A set of 120  $3'' \times 5''$  cards was used. The pronouns "I," "we," "he," "you," "she," and "they" were listed at the bottom of each card in the center of which was a verb in the simple past tense. The arrangement of the pronouns was random, and the 30 verbs used were selected from among the 200 that Dixon and Dixon (4) scaled for impression value in terms of the degree of good-to-bad impression given of the person using them. Three levels of verb-impression value were used: 10 verbs connoted a good impression; 10, a neutral impression; and 10, a bad impression [see Dixon (2)]. Each *S* was told that she was to participate in a personality study, and that *E* would learn about *S*'s personality by the way *S* talked and constructed sentences.

The task, like Taffel's [see (7)], required *S* to choose one of the six pronouns as the first word of a sentence, to use the verb as the second word, and then to complete the rest of the sentence. In each block, *S* was given 30 cards: 10 with good-impression verbs, 10 with neutral-impression verbs, and 10 with bad-impression verbs. Immediately after finishing the experiment, *S* answered an awareness questionnaire to determine if she could state the reinforcement contingency or any relationship between the verbs and the pronouns [see Dixon (2)].

### C. RESULTS

A repeated-measure factorial-design analysis of variance [see Winer (8)] was used to determine the effects of the experimental variables, which were (a) verbs at three impression levels (good-impression, neutral-impression, bad-impression), (b) two reinforcement levels (reinforcement and no reinforcement), and (c) two response-class levels ("I" and "we" vs. "he" and "they"). Neuman-Keuls sequential-range multiple-comparisons tests were used to test differences between response-class means under the various experimental conditions [see Table 1].

TABLE 1  
MEAN FREQUENCIES OF TWO PRONOUN-RESPONSE CLASSES IN THREE  
VERB CONDITIONS DURING OPERANT AND ACQUISITION PERIODS

Period and response class	Treatment					
	Good- impression verbs	Reinforced Neutral- impression verbs	Bad- impression verbs	Good- impression verbs	Not reinforced Neutral- impression verbs	Bad- impression verbs
Operant period						
I, we	8.25	6.40	3.05	8.10	8.25	2.25
he, they	5.60	7.55	10.05	6.35	7.15	9.50
Acquisition period						
I, we	9.90	8.25	4.60	7.10	7.05	2.95
he, they	7.95	9.35	10.20	7.20	6.35	9.15

Analysis of variance indicates the main effects of all three experimental variables to be significant: for verbs,  $F = 8.86$  ( $p < .01$ ); for response class,  $F = 20.03$  ( $p < .001$ ); for reinforcement,  $F = 4.74$  ( $p < .05$ ). Of particular importance, the verb-by-response-class interaction is highly significant ( $F = 101.57$ ,  $p < .001$ ). In sum, the results indicate successful conditioning and a systematic, differential effect of the good-impression, neutral-impression, and bad-impression verbs upon the two response classes.

In the two operant blocks, the choice of the personal pronouns "I" and "we" was most frequent with good-impression verbs and least frequent with



bad-impression verbs. The reverse occurred for the impersonal pronouns "he" and "they"; i.e., "he" and "they" were used most frequently with bad-impression verbs and least frequently with good-impression verbs (in all cases,  $p < .01$ ).

In the case of both response classes, neutral-impression verbs did not result in response frequencies (for "I" and "we" or "he" and "they") significantly different from those response frequencies associated with good-impression verbs; nor were significant differences found *between* the frequencies of the two response classes when they were used with neutral-impression verbs.

For reinforcement, neither the response class-by-reinforcement interaction nor the verb-by-reinforcement interaction is significant; moreover, essentially the same increase occurred for both response classes in the three verb-impression conditions. One should note, however, that reinforcement was not effective for the impersonal pronouns "he" and "they" in the bad-impression condition.

Three Ss were replaced because they were able to indicate the contingency between the reinforcing stimulus and the appropriate response, but most Ss made comments suggesting they thought E's comment "good" was simply an attempt to encourage S to continue or they did not regard it as specifically related to what S was doing. Many comments indicated concern over the fact that S *had to use* the verbs indicated—particularly the bad-impression stimuli. Some Ss stated they would not ordinarily use such words, and many Ss felt certain that the verbs were being used somehow to test their personalities; e.g., "I think you were trying to find out what I'd say about those verbs," or "you wanted to know what I think about those verbs." Further questioning, however, did not indicate any awareness of systematic and differential use of pronouns with particular verbs. Apparently the personality-test instructions and the use of various impression-value verbs were effective in providing a distracting "face validity" that minimized reportable awareness of the reinforcement contingency or the differential verb-pronoun relationship.

#### D. DISCUSSION

Our major conclusion is that previously learned verbal habits are reflected in verbal conditioning of the sentence-completion type, insofar as good-impression verbs and bad-impression verbs differentially affect the frequency of *both* the personal-pronoun response class and the impersonal-pronoun response class apart from experimental reinforcement. In general, Ss use self-referents with greater frequency when the verb-determined context connotes a good

impression; and they use self-referents with lesser frequency in a bad-impression context. The opposite is true when other referents are used.

The foregoing results extend the number of subject pronouns that have been demonstrated to be affected systematically by good-impression verbs and bad-impression verbs. It seems probable, therefore, that well-established relationships also exist between such verbs and other subject pronouns. In fact, many such previously learned intrasentence relationships may also exist between other words in sentences; and such previous learning may also affect verbal-conditioning results.

Because the neutral-impression verbs did not result in a subject-pronoun frequency that is significantly different from that obtained with good-impression verbs, and because the frequencies of the two response classes are not significantly different in the neutral-impression-verb condition, a neutral impression may be equivalent to "no impression or (stated differently) neutral-impression stimuli may not serve as stable or effective discriminative stimuli for the subject pronouns in the sentence-completion situation under study. The possibility that a neutral impression equals zero impression in these circumstances would provide a unique behavior support for the scaling assumption that the neutral point has a zero value on a continuum that runs from one extreme to another.

The reasons that "good" is equally effective in all but the he-they bad-impression-verb condition are not readily apparent. It is possible, however, that the relatively high frequency of "he" and "they" in the operant period accounts for the lack of a substantial increase in the conditioning period. In short, the response frequency may have been close to its asymptote, at least under the present conditions of reinforcement; therefore, it could not increase appreciably. These circumstances would be equivalent to a "ceiling effect," an effect that seems particularly plausible in view of the fact that none of the other response frequencies exceeded the he-they, bad-impression, operant-period frequency. In addition, as Dixon has noted, response frequencies in verbal conditioning of the sentence-completion type rarely exceed 50 per cent (the proportion also observed in the present study) of the possible frequency, unless *Ss* are clearly aware of the reinforcement contingency (in which case they occasionally reach 100 per cent).

The present data on awareness corroborate Dixon's previous findings with this procedure and materials and support the contention that the probability of *S* discerning the reinforcement contingency may be decreased by using instructions and a procedure that encourage attention to factors *other* than the



reinforcement contingency; i.e., when *S* is told that a test of personality is involved (and when what *S* does appears to be consistent with such instructions), there is little reason for *S* to attend the reinforcing stimulus or (more importantly) to the reinforcement contingency. Undoubtedly, most *Ss* are intellectually capable of solving what is essentially a very simple contingency if they attempt to do so, but the verbal-conditioning literature reporting awareness indicates that little, if any, attempt has been made to develop instructions, procedures, and materials to minimize the probability that *S* be sufficiently curious about the reasons for the occurrence of the reinforcing stimulus to try to solve the contingency involved.

The present results extend the number of pronoun responses that are known to be systematically affected (throughout a verbal-conditioning procedure) by previously established connections with verbs that connote either a good impression or a bad impression; moreover, such previously learned verbal habits are greater determinants than is experimentally controlled reinforcement of the critical-response frequencies to be observed in the conditioning experiment.

#### REFERENCES

1. BUSS, A., & DURKEE, A. Conditioning of hostile verbalizations in a situation resembling a clinical interview. *J. Consult. Psychol.*, 1958, **22**, 415-418.
2. DIXON, T. R. The effect of pre-experimentally acquired verbal habits upon verbal conditioning. *J. of Psychol.*, 1965, **59**, 335-347.
3. ———. The effect of experimenter-subject relationship upon verbal habits and verbal conditioning. *J. Gen. Psychol.*, in press.
4. DIXON, T. R., & DIXON, J. F. The impression value of verbs. *J. Verb. Learn. & Verb. Behav.*, 1964, **3**, 161-165.
5. FERGUSON, D. C., & BUSS, A. H. Operant conditioning of hostile verbs in relation to experimenter and subject characteristics. *J. Consult. Psychol.*, 1960, **24**, 324-327.
6. SIMKINS, L. Generalization effects of hostile verbal reinforcement. *J. Personal.*, 1961, **29**, 64-72.
7. TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. Abn. & Soc. Psychol.*, 1955, **51**, 496-501.
8. WINER, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Department of Psychology  
University of Kentucky  
Lexington, Kentucky 40506

## S-R RELATIONSHIPS AND TRANSFER OF TRAINING\*

*Airesearch Manufacturing Company, Division of The Garrett Corporation  
Los Angeles*

---

E. C. WORTZ AND A. C. MCTEE

---

### A. INTRODUCTION

The effect of the number of items to be learned on the rate of learning has been amply demonstrated by Ebbinghaus (2) and countless successors. In considering transfer of training, however, the rate of learning in the new situation is generally assumed to be a function of the similarity between the initial stimuli (or practice stimuli) and the test stimuli. Further, the greater the similarity between the practice stimuli and the test stimuli the greater the amount of positive transfer. As Osgood (6) states them, three laws of the effects of stimulus and response alteration on transfer of training are as follows:

When stimuli are varied and responses are functionally identical, positive transfer and retroactive facilitation are obtained, the magnitude of both increasing as the similarity among stimulus members increases.

When stimuli are functionally identical and responses are varied, negative transfer and retroactive interference are obtained, the magnitude of both decreasing as similarity between the response increases.

When both stimulus and response members are simultaneously varied, negative transfer and retroactive interference are obtained, the magnitude of both increasing as stimulus similarity increases (6, pp. 526-529).

The general concept of the relationship between training-test similarity and interference with transfer is that the greater the similarity the greater the interference. The paradox of similarity between current events and preceding events being simultaneously the condition for both maximal facilitation of and interference with learning led to the Skaggs-Robinson hypothesis. Robinson (7) hypothesized a curvilinear relationship between similarity and facilitation, with greatest facilitation at maximum similarity and minimum similarity.

Osgood (5) has proposed a complex transfer-and-retroaction surface that (in a qualitative manner) accounted for all previous data, resolved the paradox, and is consistent with his laws of transfer of training.

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on July 8, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



One common paradigm of transfer-of-training situations that has been subjected to little experimental scrutiny involves the use of identical stimuli and responses for both the training and the test situations. In this paradigm, stimulus and response relationships are learned during the initial period; and new relationships between the same stimuli and responses must be learned during the test period. This paradigm does not have an appropriate place in Osgood's transfer-and-retroaction surface or on Houston's modified version (3) of that surface.

A molar view that provides predictive possibilities is offered by the discrimination hypothesis formulated by Mowrer (4). It follows from the discrimination hypothesis, that, as training and reversal periods become increasingly dissimilar, the rate of extinction of the initial responses increases. Rate of transfer is then facilitated by features that emphasize the dissimilarity between practice periods and test periods. Alteration of reinforcement schedule defines, by definition for this paradigm, the extent of difference between the practice and the test period. In addition, at least one other factor (that of the number of new S-R relationships) should operate here. The parsimonious prediction of the effect of the number of elements changed is supplied by Ebbinghaus, as discussed earlier.

In a transfer-of-training situation involving only S-R relationships, without the changing of stimuli or responses, transfer of training should be more rapid the greater the alteration in reinforcement (S-R relationships changed) and the fewer the number of items altered. These two techniques of facilitating transfer however, are, antithetic; consequently, the effect of one should nullify the effect of the other, the combined effects should "peak," or one should evidence a greater effect than the other. Difficulty of transfer, then, could be a horizontal, a curvilinear, or a monotonic function of the number of S-R relationships altered.

The purpose of this experiment, then, is to ascertain the effect of training and test similarity on difficulty of transfer for the paradigm in which no new stimuli or responses are employed in the test period. Only simple rearrangement of previously learned S-R relationships is provided.

## B. METHOD

### 1. Apparatus

The experimental apparatus consisted of:

1. A stimulus panel, comprising six orange lights, each one-half inch in diameter, and one translucent white light one inch in diameter. The orange lights—spaced one-half inch apart—were arranged in a matrix of three rows

and two columns; and the white light was located midway between the columns, three inches below the bottom row. All lights were mounted in a plywood panel painted flat black. A double-pole, double-throw toggle switch was provided for the subject's response.

2. An experimenter's panel, comprised of switches for controlling the individual lights in the matrices (for turning on the stimulus patterns) and a small neon lamp that indicated correct responses.

3. An electronic interval timer, which controlled the duration of the correct-response light.

4. A hand-tally counter for recording correct responses.

5. A randomizing device (manufactured by the Transogram Company).

The subject was seated in a chair that roughly simulated an aircraft seat. The stimulus panel was placed vertically, approximately 28 inches in front of the subject's normal eye position, with the center of the matrix approximately 27 inches above the seat surface. The subject's response switch was mounted on a horizontal shelf below the stimulus panel, with the handle vertically upward. Actuation was either left or right.

## 2. Subjects

Thirty-five adult males were selected randomly from an engineering department to serve as subjects, and each subject was randomly assigned to an experimental group.

## 3. Procedures

The subjects were required to learn the correct switch position (left or right) in response to a pattern of two lights in a random series of all possible (i.e., 15) two-light patterns in a three-by-two matrix of orange lights. The sequence of light patterns was randomized for each trial. Correct responses were rewarded by a white light situated separately from the matrix.

When *S* was seated, *E* read the following set of instructions:

The purpose of this experiment is to determine the efficiency of the human operator in a signal-detection task. In front of you is a matrix or orange lights (*E* turns on all lights in the matrix) and one large white light. A pattern of two lights, similar to this one (*E* turns off all lights except those in the top row), will appear in the matrix. You will respond to the lights by pushing your switch either left or right. If your choice is correct, the white light will come on for two seconds; if your choice is incorrect, the white light will not come on. When the lights in the matrix go off, return your switch to the center position and wait for the next pattern.



Now push your switch to the left. The white light will come on momentarily. (*E* actuates the correct-response switch to give *S* the correct-response light.) Now push your switch to the right. The white light should come on again. Now return your switch to the center. (*E* turns off all lights.)

Your task is to move your switch either left or right, so that the white light will come on every trial; in other words, you are to learn the correct switch position for each light pattern.

Do you have any questions? (Questions were answered by repeating or rehearsing the instructions.) Ready?

One presentation of all 15 light patterns was considered as one trial. The criterion of learning was two successive errorless trials.

After reaching the criterion, the correct response was reversed for a certain percentage of the light patterns. The number of patterns for which responses were reversed determined the experimental groups. Each subject was required to learn to a criterion of 100 per cent correct on each of two consecutive trials; the correct responses for the 15 patterns; then the correct response for either 20, 40, 60, 80, or 100 per cent (three, six, nine, 12, or all 15) of the patterns was reversed, and *S* relearned to the same criterion—15 correct on each of two consecutive trials. The individual light patterns for which the responses were reversed were randomly selected for all *Ss* in each group. The ratio of the number of trials to criterion following response reversal and the number of trials to the learning criterion ( $NTR/NTL$ ) was computed for each individual and used as the index of transfer difficulty.

### C. RESULTS

An analysis of variance, summarized in the first two substantive columns of Table 1, shows no difference between the groups during the training trials.

Difficulty of transfer (response reversal), as indicated by the mean ratio

TABLE 1  
SUMMARY OF ANALYSES OF VARIANCE

Source	<i>df</i>	Prereversal trials*		For the ratio $NTR/NTL^{1,2}$	
		Sum of squares	Mean square	Sum of squares	Mean square
Treatments	4	43.88	10.97	2.49	.62
Within	30	1790.87	59.69	4.10	.14
Total	34	1834.75		6.59	

\*  $F = 0.18$  (not significant).

<sup>1</sup> The number of trials to criterion following response reversal divided by the number of trials to the learning criterion.

<sup>2</sup>  $F = 4.58$ ,  $p < .01$ .

of the number of trials to relearn and the number of trials for initial learning—when computed on an individual basis—tends to eliminate individual differences; and an analysis of variance (summarized in the last two columns of Table 1) indicates that it is unlikely that the differences among groups (.48, .92, 1.12, .64, and .46—as 20 per cent, 40 per cent, 60 per cent, 80 per cent, and 100 per cent of the patterns were reversed) arises by chance.

Table 2 shows the *t* scores between groups. Transfer for the 20 per cent reversal group is significantly faster than that for the 40 per cent group or

TABLE 2  
SIGNIFICANCE TESTS (*t*) BETWEEN GROUPS FOR THE RATIO NTR/NTL\*

Treatment	20	Treatment 40	60	80
40	2.45**			
60	2.90**	.056		
80	0.75	1.36	1.81	
100	0.14	2.55**	3.00**	0.85

\* The number of trials to criterion following response reversal divided by the number of trials to the learning criterion.

\*\*  $p < .05$ .

the 60 per cent group, but is not significantly different from that for the 80 per cent group or 100 per cent group. The 100 per cent reversal group, also, is significantly faster than the 40 per cent group or the 60 per cent group. Consequently, a curvilinear function between transfer (reversal) difficulty and the number of S-R relationships altered is established.

In general these data fit the basic form of the Skaggs-Robinson hypothesis. The curvilinear relationship shown for transfer difficulty as a function of training-test similarity is beyond the predictive range of the more-rigorous learning theories; also, this relationship is contradictory to Osgood's third law of transfer of training and is incompatible with his transfer-and-retroaction surface. This experiment, however, involves a new dimension for training-test similarity experiments; namely, that of variations in training-test similarity without variation in stimuli or responses.

It is uncertain, therefore, whether or not the experiment represents a special case of transfer of training; or whether or not Osgood's interpretation of the effects of similarity for the special case of paired-associates learning is incorrect. The authors agree with Bugelski and Cadwallader (1) that much of the argument on the effects of similarity on transfer depend upon the definition of similarity.

In considering the utility of the discrimination hypothesis, these results appear to be in general accord with the concept but not directly predictable from



it. This type of experiment, however, lends itself to the further exploration and possible quantification of the operations and mechanisms implied by the discrimination hypothesis.

## REFERENCES

1. BUGELSKI, B. R., & CADWALLADER, T. C. A reappraisal of the transfer and retroaction surface. *J. Exper. Psychol.*, 1956, **52**, 360-366.
2. Ebbinghaus, H. Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Leipzig: Dunker & Humboldt, 1885 (translated by H. A. Ruger and C. E. Bussenius as *Memory: A Contribution to Experimental Psychology*. No. 3. *Coll. Educ. Reprints*, Teachers College, Columbia University, 1913.
3. HOUSTON, J. P. Verbal transfer and interlist similarities. *Psychol. Rev.*, 1964, **71**, 412-414.
4. MOWRER, O. H., & JONES, H. Habit strength as a function of pattern of reinforcement. *J. Exper. Psychol.*, 1945, **35**, 293-311.
5. OSGOOD, C. E. The similarity paradox in human learning. *Psychol. Rev.*, 1949, **56**, 134-154.
6. ———. *Method and Theory in Experimental Psychology*. New York: Oxford, 1953. Pp. 526-529.
7. ROBINSON, E. S. The similarity factor in retroaction. *Amer. J. Psychol.*, 1927, **39**, 297-312.

Airesearch Manufacturing Company  
9851-9951 Sepulveda Boulevard  
Los Angeles, California 90009

The Bunker-Ramo Corporation  
Los Angeles, California

## LOGICAL THINKING AND PARANOID SCHIZOPHRENIA\*<sup>1</sup>

*Milledgeville State Hospital, Georgia*

F. A. COYLE, JR. AND J. L. BERNARD

### A. INTRODUCTION

Inherent in the concept of<sup>2</sup> the schizophrenic disorders is a disturbance in cognition. It is commonly assumed that this disturbance results in bizarre thinking and behavior and in a general intellectual deficit. Prominent in psychiatric circles is Silvano Arieti's conception regarding schizophrenic thought disorder; i.e., that the disturbance of the schizophrenic is one of paleological (prearchaic, prelogical, paralogical) thinking. Arieti (2) extended the Von Domarus model (18) of the laws governing schizophrenic thought. According to Von Domarus, the schizophrenic accepts the identity of predicates as a basis for valid conclusions, as opposed to the views of normals whose syllogistic reasoning is based on the Mode of Barbara (or identical subjects).

Recent investigators have attempted to assess the validity of the Arieti-Von Domarus hypothesis; and, being mindful of an earlier study (22) that indicated that the difficulty of a syllogism is affected by the wording, have attempted to control for this. Moore and Anderson (11) developed a series of steps based on the calculus of propositions and hoped that future investigators would use these in human problem-solving experimentation. They felt the complexity of a task could be controlled and standardized, and that equivalent tasks could be developed; however, few (if any) investigators have followed Moore and Anderson's lead.

Gottesman and Chapman (8) made their material available through the American Documentation Institute; but it lacks standardization in development and, realistically, has little potential for widespread use.

Recent studies are valuable, however, in that they are consistent in their findings. All of them lack experimental evidence for the Arieti-Von Domarus hypothesis. Chapman and Chapman (5), for example, found no support for the hypothesis that a particular type of syllogistic reasoning error is an im-

\* Received in the Editorial Office, Provincetown, Massachusetts, on July 8, 1965, and published immediately at 35 New Street, Worcester, Massachusetts, Copyright by The Journal Press.

<sup>1</sup> This article is based on a paper presented at the Southeastern Psychological Association, Atlanta, Georgia, April 2, 1965.



portant aspect of schizophrenic thought, as is argued by Von Domarus. Gottesman and Chapman (8), using 30 schizophrenics and 30 normals, found no significant difference on the Von Domarus items and concluded that their findings did not support Von Domarus' or Arieti's description of a particular type of schizophrenic error. (It is pertinent to the present study to note that all their schizophrenics had a chronic diagnosis.)

J. P. Nims (12) studied 25 hospitalized male schizophrenics and 25 "normal" controls from a TB hospital. He does not provide a breakdown of specific diagnoses or symptomatology in his schizophrenic group, but (as in the other studies mentioned) tested them on an instrument of his own construction. Like other investigators, Nims found no difference between the groups. Interestingly, Nims included emotionally "charged" items; yet (contrary to his hypothesis) there was no difference in the performance of schizophrenics and normals on items of this type.

Williams (20) tested 50 hospitalized psychiatrically normal TB patients and 50 patients manifesting various types of schizophrenia. He employed four content categories (abstract, concrete, personal, and nonpersonal), but he found no significant difference in any category.

Williams (20) and Nims (12) have concluded that it is not the type of logic employed by the schizophrenic (but other factors) that differentiates them from normals. Williams has pointed to the fact that Arieti assumes that the normal person automatically applies Aristotelian laws of logic, and it is at this crucial point that Williams and Arieti disagree. Williams contends that "... deductive reasoning plays a small part in the thinking of schizophrenics. Indeed, normals do not seem to rely upon it excessively" (20, p. 60). Further, Williams hypothesizes that most thinking is basically irrational. In addition, after an exhaustive review, Payne (14) concluded that "... there is no evidence that any abnormal group have a specific disability for tests involving deductive reasoning" (14, p. 223).

To summarize, all the studies cited have two weaknesses: i.e., the use of "homemade" tests and the use of a schizophrenic group of mixed or unstated diagnosis. In spite of these weaknesses, however, all of the studies indicate that on logic tasks or reasoning tasks schizophrenics do not differ from normals.

In addition to the assumption of a basic schizophrenic thought disorder, there is a diversity of opinion about the logic of the paranoid schizophrenic. As with many issues, there are three camps: pro, con, and ambivalent. A sampling follows:

1. *Pro*

a. *Harry Stack Sullivan*. "... it is only the individual who develops somewhat logical beliefs in persecution . . . that comes under the unfavorable rubrics of the paranoid group" (17, p. 66).

b. *Gregory A. Kimble*. "... given the paranoid's basic assumption, the delusory system is developed in a perfectly logical manner. . . . This contrasts so sharply with the illogicality and rambling incoherence of the other schizophrenic groups that there is some question as to whether the paranoid is properly included in the schizophrenic category" (9, p. 383).

c. *Ewalt, Strecker, and Ebaugh*. "... and, typically, the patients have reasonable well-organized delusional systems, which always contain some imagined harm or insult to the patient" (7, p. 216).

2. *Con*

a. *Coleman*. "... the eventual symptom picture is dominated by absurd, illogical, and changeable delusions" (6, p. 251).

b. *Morgan*. "Unlike the pure paranoid, the paranoid schizophrenic has unsystematized delusions—delusions that are not too coherent" (10, p. 70).

c. *Smith and Smith*. "... the delusions of paranoid schizophrenia are not logically organized . . ." (16, p. 467).

d. *Noyes and Kolb*. "... the features that tend to be more evident in this type . . . or phase are delusions, which are often numerous, illogical, and disregarding of reality. . . ." (13, p. 347).

e. *Cameron*. "... the delusions are as a rule persecutory in character, but may also be grandiose, ecstatic, mystical or hypochondriacal; they are never as well-organized and systematized as are the delusions of classical paranoia" (4, p. 448).

3. *Ambivalent*

a. *Silverman*. "The delusional system may be highly elaborated and systematized, relatively unsystematized, or over time it may become less systematized" (15, p. 354).

b. *Arieti*. "Although the delusions are unsystematized in the majority of cases, they are more systematized at the beginning of the illness" (3, p. 461).

## B. THE PRESENT STUDY

1. *Subjects*

Subjects<sup>2</sup> were selected from the patient population at Milledgeville (Georgia) State Hospital. The basic criteria for selection were that the

<sup>2</sup> The authors would like to thank S. Donehoo and P. Weaver for their assistance in gathering data.



patient (*a*) entered the hospital after January 1, 1960—the attempt was to minimize chronicity, (*b*) had a minimum of 10 years of schooling, (*c*) was between the age of 21 and 51, and (*d*) was Caucasoid. Both males and females were selected from three diagnostic groups: schizophrenic reaction, paranoid type; schizophrenic reaction, chronic undifferentiated type; and psychoneurotic disorder (mixed types).

The authors felt that the two schizophrenic groups were sufficiently dissimilar in terms of symptomatology, prognosis, and reality contact to demonstrate a difference (if one existed) in logical ability. The psychoneurotics were chosen as a control group because the authors felt that (while psychoneurotics evidence none of the usual schizophrenic thought disorder or reality distortion), they are institutionalized individuals with present environmental factors and socioeconomic backgrounds similar to those of the two schizophrenic groups.

The mean age of each group was approximately 35, and the mean number of school grades completed was 11. The number of cases in each group was as follows: schizophrenic reaction, paranoid type, 32 (18 males, 14 females); schizophrenic reaction, chronic undifferentiated type, 26 (13 males, 13 females); and neurotics, 26 (10 males, 16 females).

We included all the paranoid schizophrenics in the hospital (which has 12,000 beds) who met our criteria, with the exception of three who refused to be tested. Originally, we planned to test patients exhibiting paranoid states and paranoid personalities, but only three were found in the hospital. All patients were selected by means of an IBM filing unit that furnished a random selection of chronic undifferentiated schizophrenics and psychoneurotics.

## 2. Tests

The principal test given to the subjects (in small groups) was the Watson-Glaser Critical Thinking Appraisal (19), which has five parts as follows: Inference (20 items), Recognition of assumptions (16 items), Deduction (25 items), Interpretation (24 items), and Evaluation of arguments (15 items).

To control for correlation between intelligence and performance on the Watson-Glaser test, all subjects were given Form 1 of Ammons and Ammons' Quick Picture Vocabulary Test (1), but no group or sex difference was found to be significant.

## C. RESULTS

The scores on the Watson-Glaser test were subjected to a two-by-three analysis of variance, with unequal cell frequencies (21). An *F* value of 1.03

was obtained for the sex variable; one of 1.10 for the diagnosis variable; and one of .64 for the interaction effect. From these values (no one of which is significant), it appears that among the groups tested there was no difference in the ability to think logically and critically. On the basis of the Watson-Glaser test, paranoid schizophrenics do not seem to be superior or inferior to chronic undifferentiated schizophrenics or neurotics in critical ability.

#### D. DISCUSSION AND SUMMARY

From the results of earlier studies, the authors concluded that there are no particular reasoning or logical difficulties that beset schizophrenics that do not also affect normals to an equal degree; yet there remains considerable disagreement among theoreticians and practitioners about the logical reasoning abilities of paranoid schizophrenics. This study indicates that paranoid schizophrenics are no different in their thinking from chronic undifferentiated schizophrenics or neurotics. Despite this, Moore and Anderson (11) have pointed out that "To identify 'laws of logic' with 'laws of thought,' or, in general, to confuse the logical aspects of problems and the psychological aspects of problem solving is a mistake not infrequently made" (11, p. 156). Perhaps, then, the schizophrenic (in general) and the paranoid schizophrenic (in particular) are as adept (or as inept) as are normals in the use of "laws of logic."

It may be that the logic used by most people to justify behavior is illogical at times; however, it is hypothesized that the paranoid schizophrenic makes repeated appeals to "logic" (as well as to "truth," "justice," and "fairplay") and that this serves to point out the logical deficit in the patient's thought. This deficit would be apparent in anyone else if he called attention to it by appealing to "logic" as a source of support for belief and behavior.

Possibly, ability to think logically cannot be measured by paper-and-pencil tests; yet it seems reasonable to assume that psychometric methods can be applied to this area as well as to the measurement of intelligence, achievement, and personality. If the applicability of testing is granted, one can argue that schizophrenics are capable of employing logic in an impersonal testing situation but not in an interpersonal setting. The inclusion of items with emotional overtones attempted to cover this eventuality; however, it is obvious that the emotionality and logicity employed by schizophrenics cannot be measured directly by a test like the Watson-Glaser Critical Thinking Appraisal.

The psychoneurotics tested in the present study are a particularly appropriate control group. They were equivalent in age, socioeconomic background,



and intelligence to the schizophrenics. In addition, they were in the same general hospital environment as were the schizophrenics. The similarity of present environments is greater than in studies using tuberculosis-hospital patients or penal-institution inmates as controls.

Because earlier studies indicate no difference between normals and schizophrenics in logical ability, the present study implies that neurotics, normals, and schizophrenics are not different in the use of logical processes; therefore, the authors conclude that chronic undifferentiated schizophrenics, paranoid schizophrenics, psychoneurotics, and normals do not differ in their thinking processes. Perhaps Williams (20) is correct in asserting that in daily thinking there is little application of logic by anyone.

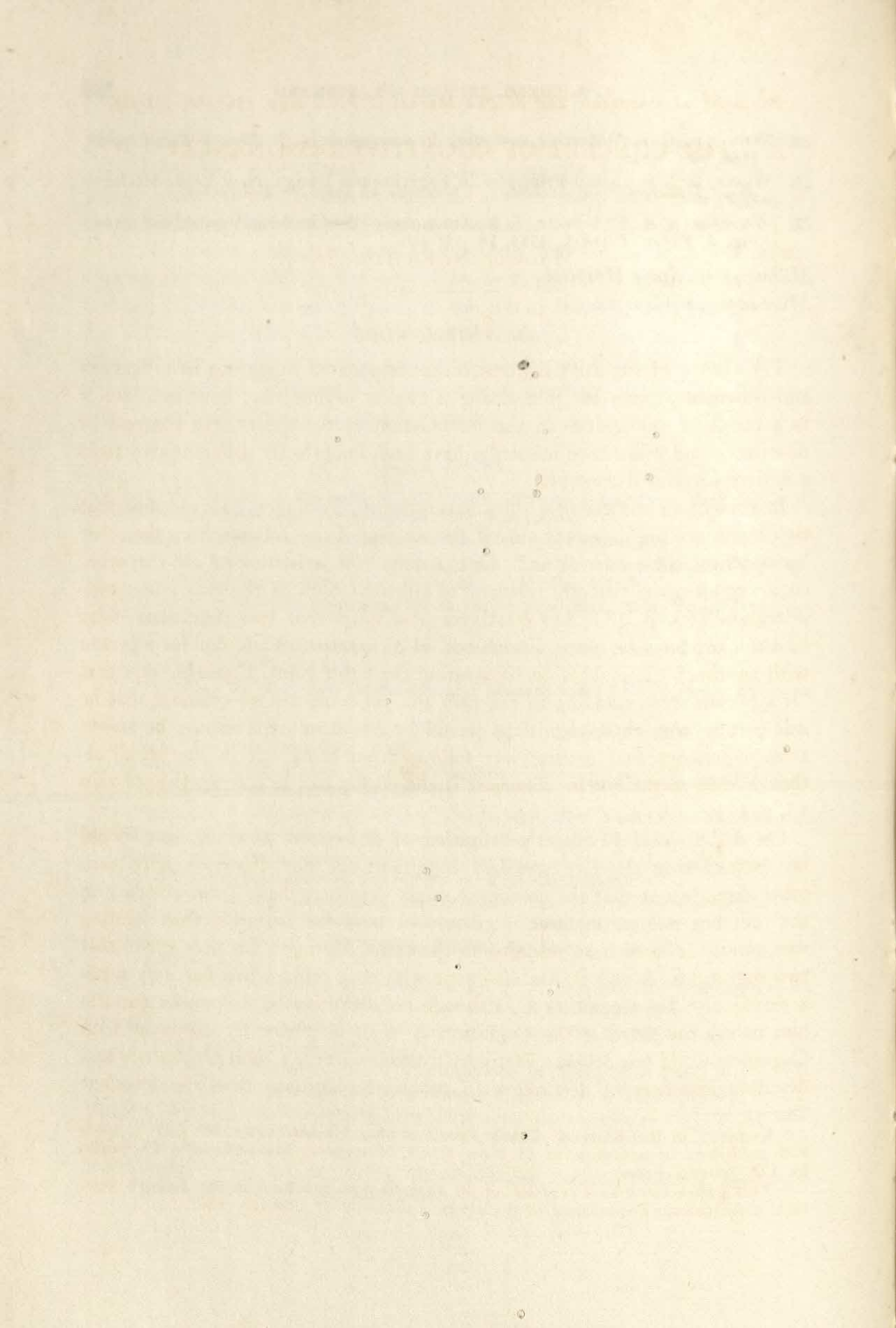
#### REFERENCES

1. AMMONS, R. B., & AMMONS, C. H. Quick Picture Vocabulary Test: Form 1. Missoula, Montana: Psychological Test Specialists, 1958.
2. ARIETI, S. Interpretation of Schizophrenia. New York: Brunner, 1955.
3. ———. Schizophrenia. In S. Arieti (Ed.), *The American Handbook of Psychiatry*. Volume 1. New York: Basic Bks, 1959. Pp. 445-484.
4. CAMERON, N. A. The Psychology of Behavior Disorders. New York: Houghton Mifflin, 1947.
5. CHAPMAN, L. J., & CHAPMAN, J. P. Atmosphere effect re-examined. *J. Exper. Psychol.*, 1959, **58**, 220-226.
6. COLEMAN, J. C. Abnormal Psychology and Modern Life. New York: Scott-Foresman, 1950.
7. EWALT, J. R., STRECKER, E. A., & EBAUGH, F. G. Clinical Psychiatry. (8th ed.). New York: McGraw-Hill, 1957.
8. GOTTESMAN, H., & CHAPMAN, L. J. Syllogistic reasoning errors in schizophrenia. *J. Consult. Psychol.*, 1960, **24**, 250-255.
9. KIMBLE, G. A. Principles of General Psychology. New York: Ronald, 1956.
10. MORGAN, C. T. Introduction to Psychology. New York: McGraw-Hill, 1956.
11. MOORE, O. K., & ANDERSON, S. B. Modern logic and tasks for experiments on problem solving behavior. *J. of Psychol.*, 1954, **38**, 151-160.
12. NIMS, J. P. Logical reasoning in schizophrenia: The Von Domarus principle. Unpublished Doctoral dissertation. University of Southern California, 1959.
13. NOYES, A. P., & KOLB, L. C. Modern Clinical Psychiatry. (6th ed.). Philadelphia: Saunders, 1963.
14. PAYNE, R. W. Cognitive abnormalities. In H. F. Eysenck (Ed.), *Handbook of Abnormal Psychology*. New York: Basic Bks, 1961. Pp. 193-261.
15. SILVERMAN, J. The problems of attention in research and theory in schizophrenia. *Psychol. Rev.*, 1964, **5**, 352-379.
16. SMITH, K. W., & SMITH, W. U. The Behavior of Man. New York: Holt, 1958.
17. SULLIVAN, H. S. Schizophrenia as a Human Process. New York: Norton, 1962.
18. VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasonin (Ed.), *Language and Thought in Schizophrenia*. Berkeley: Univ. Calif. Press, 1946. Pp. 104-114.
19. WATSON, G., & GLASER, E. M. Watson-Glaser Critical Thinking Appraisal: Manual. New York: Harcourt, 1964.

20. WILLIAMS, E. B. Deductive reasoning in schizophrenia. *J. Abn. & Soc. Psychol.*, 1964, **69**, 47-61.
21. WINER, B. J. *Statistical Principles in Experimental Design*. New York: McGraw-Hill, 1962.
22. WOODWORTH, R. S., & SELLS, S. B. Atmosphere effect in formal syllogistic reasoning. *J. Exper. Psychol.*, 1935, **18**, 451-460.

*Milledgeville State Hospital*  
*Milledgeville, Georgia*





## THE CONCEPT OF COGNITIVE DISSONANCE\*<sup>1</sup>

*Department of Sociology, University of Southern California*

---

MAURICE N. RICHTER, JR.

---

### A. INTRODUCTION

The theory of cognitive dissonance has stimulated numerous investigations and provided a basis for integrating a variety of findings; however, largely as a result of ambiguities in the formulation of the theory, the concepts of dissonance and dissonance reduction have been imperfectly differentiated from a variety of related concepts.

In presenting the cognitive-dissonance theory, Festinger (3) specified that two cognitions are dissonant only if the obverse of one follows from the other "considering these two alone," disregarding "the existence of all the other cognitive elements that are relevant to either or both of the two under consideration" (3, p. 13); but Festinger also states that two cognitions "may be dissonant for a person . . . with one set of experiences and not for a person with another" (3, p. 15). In illustrating the latter point, Festinger says that "if a person were standing in the rain and yet could see no evidence that he was getting wet, these cognitions would be dissonant . . . because he knows from experience that getting wet follows from being out in the rain," although they might not be dissonant if the person had never experienced rain (3, p. 14).

On the basis of Festinger's definition of dissonance, however, one should say instead that the two specified cognitions are not dissonant with each other directly but that the combined single cognition that "I am standing in the rain but not getting wet" is dissonant with the cognition that "getting wet usually follows from standing in the rain." Festinger has thus stated that two cognitions, A and B, are dissonant with each other when *but only when* a person also has Cognition C, although his definition of dissonance requires him to say instead that the Cognition  $A + B$  is inherently dissonant with Cognition C. If one follows Festinger's definition rather than his practice and describes instances of dissonance in terms of cognitions that are dissonant

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on July 9, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

<sup>1</sup> This paper contains a revision of an analysis that appeared in the author's Doctoral dissertation, Department of Sociology, University of Chicago, 1962.



inherently rather than merely in particular contexts, the possibility of analyzing the dissonance concept on a logical level, as well as on a psychological level, arises.

### B. ANALYSIS

Stating that a person has a particular cognition or belief may be considered equivalent to stating that he accepts a particular proposition. On this basis, one can distinguish between *cognitive* dissonance as a relationship between cognitions and *logical* dissonance as a relationship between propositions. Specifically, logical dissonance may be recognized as the kind of relationship that exists between the propositions that a person is accepting when cognitive dissonance occurs.

As Osgood (5) has pointed out, the tension postulated to exist when a person has dissonant cognitions can occur only if the cognitions interact. Logical dissonance can exist, therefore, only between propositions that a person can accept simultaneously under conditions of interaction: i.e., only between propositions that are logically compatible with each other in terms of the person's own premises.

Festinger claims that cognitive dissonance *sometimes* involves logical contradiction but *usually* involves propositions that tend to negate each other only psychologically. The present author's conclusion is that there is *never* a genuine logical contradiction between the propositions involved in cognitive dissonance, but that there is *always* a negative logical relationship of a milder sort between the propositions. This relationship of "logical dissonance" exists between two propositions when one proposition is a generalization and the other is an exception to the generalization.

Concepts referring to disharmonies among cognitions within an individual's belief system include not only dissonance but also imbalance, belief inconsistency of the kind that McGuire (4) has investigated, and conflict. Cognitive imbalance, as defined by Rosenberg (6), is present when a "negative relation exists between two affectively negative objects or between two affectively positive objects, or a positive relation exists between an affectively positive object and an affectively negative object" (6, p. 60). Brown has stated that imbalance in this sense is one kind of dissonance, and he assumes that dissonance does not always involve imbalance (1, p. 596). However, any case of dissonance can be interpreted as a case of imbalance by identifying the dissonant propositions as "objects," the acceptance of these propositions as a form of positive affectivity, and the dissonance between the propositions as

a "negative relationship." Instead of saying that imbalance is a variety of dissonance, one should thus recognize dissonance as a variety of imbalance.

As a basis for clarifying the relationships between dissonance and the other concepts mentioned, one can make distinctions as follows: (a) between accepting a proposition and considering its acceptance, (b) between propositions that contradict each other and propositions that are merely logically dissonant, and (c) between cognitions that interact and those that are compartmentalized. The acceptance of contradictory propositions under conditions of interaction is a logically empty category. Cognitive dissonance involves the acceptance with interaction of propositions that are logically dissonant as distinct from logically contradictory. Conflict involves the consideration, as distinct from the acceptance, of logically contradictory propositions—with the relevant cognitions interacting. Belief inconsistency, as McGuire (4) has used the concept, involves the acceptance of logically contradictory propositions—but with compartmentalization rather than interaction among the corresponding conditions. This set of distinctions is important because reduction of dissonance, reduction of conflict, and reduction of inconsistency are separate processes that sometimes coincide but that also sometimes interfere with each other.

In describing the dissonance-reducing process, one should allow for two possibilities that sometimes have been overlooked. First, although resistance to belief change is sometimes "distorting," it may sometimes have the opposite effect (as illustrated by numerous episodes in the history of science: i.e., some investigators who have remained faithful to established theories despite apparently serious anomalies have been ultimately vindicated). Second, although a person may reduce dissonance by refusing to consider evidence against an established belief, he may reduce it also by abandoning his belief on the basis of evidence against it; and resistance to belief change may sometimes reflect tolerance for dissonance rather than a successful avoidance of dissonance-creating evidence against the belief. Failure to make sufficient allowance for the foregoing possibilities has led to a partial, incomplete view of the implications of dissonance theory. Thus Festinger has suggested that a person who is relatively intolerant of dissonance has "very positive and one-sided opinions about many issues" together with an inability to make decisions (3, p. 269). Similarly, Deutsch, Krauss, and Rosenau (2) have interpreted dissonance theory as predicting that a person must "maladapt to his environment": specifically, "the more evidence a person gets that his decision . . . [is] wrong, the more ardently will he support his original decision" until he revokes it



belatedly (2, p. 17). With equal plausibility, one could predict, however, that a person who is intolerant of dissonance avoids having any opinion on any issue except when necessary and tends to abandon his beliefs whenever he finds evidence against them.

### C. CONCLUSION

It is paradoxical that an unwarranted allowance for the extreme irrationality of dissonance involving logical contradiction has been combined in the presentation of the theory with an unwarranted description of intolerance for dissonance as inherently irrational. The first of these features involves an imperfect differentiation among the concepts of dissonance, conflict, and belief inconsistency; while the second involves a failure to differentiate clearly among dissonance reduction, resistance to cognitive change, and cognitive distortion.

### REFERENCES

1. BROWN, W. *Social Psychology*. New York: Free Press, 1965.
2. DEUTSCH, M., KRAUSS, N., & ROSENAU, N. Dissonance or defensiveness? *J. Personal.*, 1962, **30**, 1-28.
3. FESTINGER, L. *A Theory of Cognitive Dissonance*. Evanston, Ill.: Row, Peterson, 1957.
4. MCGUIRE, W. A syllogistic analysis of cognitive relationships. In M. Rosenberg and C. I. Hovland (Eds.), *Attitude Organization and Change*. New Haven: Yale Univ. Press, 1960.
5. OSGOOD, C. Cognitive dynamics in the conduct of human affairs. *Pub. Opin. Quart.*, 1960, **24**, 341-365.
6. ROSENBERG, M. An analysis of affective-cognitive consistency. In M. Rosenberg and C. I. Hovland (Eds.), *Attitude Organization and Change*. New Haven: Yale Univ. Press, 1960.

6837 North Second Street  
Phoenix, Arizona 85012

## VISUAL-CHARACTER DETECTION\*

*Sylvania Electronic Systems, Waltham, Massachusetts*

---

D. W. BROWN AND H. D. FRIEDMAN

---

### A. INTRODUCTION

A specific application of work in pattern recognition is the automation of the task of the photo-interpreter. One approach is to describe the functions of the human photo-interpreter and to build a device that performs those functions. The research to be reported is a first attempt to describe the performance of subjects faced with a detection task.

It seems likely that the task of the photo-interpreter is one of detecting an unusual structure against a reasonably well-defined background. Therefore our purpose is to determine if human observers use different methods of scanning and, if so, to discover what method or methods are superior. An additional aim is to define a relatively simple laboratory task that requires some of the skills employed in photo interpretation.

The task was that of identifying a single "C" in a 16-by-16 matrix of "Gs" [see Figure 1]. The authors reasoned that the detection time for "Cs" in different quadrants would vary as a function of the scanning method employed; thus, subjects scanning from left to right and from top to bottom should find "Cs" in the upper-left quadrant fastest and take the longest time for stimuli in the lower-right quadrant. Further, the authors reasoned that the mean (or median) search times for the four quadrants would reveal the scanning pattern that each *S* employed.

Two studies were conducted. The first employed 10 subjects, with 16 matrices each, and was used to select subjects for a second study. Analysis of variance revealed a significant difference between the performance of the best five subjects and that of the remaining five. Four of the best five subjects were employed in the second experiment (the fifth *S* was unavailable).

### B. METHOD

#### 1. Apparatus

Sixteen 16-by-16 matrices were typed on an IBM electric typewriter. Each matrix contained 255 "Gs" and one "C." Placement of the "C" was essential-

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on July 9, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G

FIGURE 1  
THE STIMULUS MATRIX

ly random; but each eight-by-eight quadrant of the matrix was chosen exactly four times, and the outer borders of the quadrants were never used. A stopwatch was used to record the subject's detection times. "Cs" and "Gs" were selected as presenting a relatively difficult task and so providing a greater variation in time scores than would, for example, a "D" in a matrix of "Gs."

## 2. Subjects

For Study 1, subjects were 10 employees of the Applied Research Laboratory; for Study 2, the subjects were four of the five subjects with the fastest detection times in Study 1.

## 3. Procedure

Subjects were instructed to find the single "C" in the matrix of "Gs" and to find it as quickly as possible. All 16 matrices were presented in random order in a single session. The time from the initiation of search to the point at which the subject said he had found the "C" was recorded. The subject was required to point to the "C" on each trial.

## C. RESULTS AND DISCUSSION

## 1. Study 1

Inspection of the mean detection times for the four quadrants and the 10 subjects shows that eight scanning sequences were employed [see Table 1].

TABLE 1  
MEAN DETECTION TIMES (IN SECONDS) FOR THE FOUR QUADRANTS  
OF THE MATRIX

Quadrant	Method*										Total
	1	2	3	4	5	6	7	8	9	10	
Upper left	22.4	9.1	21.9	17.8	31.6	15.7	35.5	6.1	21.2	38.0	21.9
Upper right	40.9	20.3	45.2	12.3	40.1	14.5	32.1	23.0	6.4	11.5	24.7
Lower left	35.4	20.0	26.6	31.1	15.9	12.5	50.3	12.3	18.3	26.6	24.9
Lower right	27.6	29.8	20.6	6.4	34.4	15.8	28.3	17.2	25.8	14.8	22.1

\* The 10 methods are defined by the order in which the four quadrants were scanned: Method 1—upper left, lower right, lower left, upper right; Method 2—upper left, lower left, upper right, lower right; Method 3—lower right, upper left, lower left, upper right; Method 4—lower right, upper right, upper left, lower left; Method 5—lower left, upper left, lower right, upper right; Method 6—lower left, upper right, upper left, lower right; Method 7—lower right, upper right, upper left, lower left; Method 8—upper left, lower left, lower right, upper right; Method 9—upper right, lower right, lower left, upper left; Method 10—upper right, lower right, lower left, upper left.

The results of an analysis of variance are presented in Table 2. As can be seen, only one of the *F* ratios reaches significance.

The observation that different subjects approach the task in different ways is substantiated by the significance of the horizontal-by-vertical-by-subjects interaction (Table 2). In other words, performance on a particular quadrant (any combination of horizontal and vertical) is a function of the particular subject. A surprising finding is the lack of a significant difference among

TABLE 2  
ANALYSIS OF VARIANCE FOR THE DETECTION-TIME DATA

Source	df	Mean square	F
Horizontal (left-right)	1	0	—
Vertical (up-down)	1	.78	—
Subjects	9	1,003.25	1.04
Interaction	1		
Horizontal-by-vertical	1	316.07	0.33
Horizontal-by-subjects	9	1,567.49	1.63
Vertical-by-subjects	9	1,198.96	1.24
Horizontal-by-vertical-by subjects	9	2,127.61	2.21*
Within cells	120	964.79	
Total	159		

\* Significant at the .05 level.



subjects, but this lack is attributable to the rather large variability of subjects within cells (within-cells or error term). This conclusion is in contrast to the lack of differences either as a function of horizontal location (left half or right half of the matrix) or vertical location (upper half or lower half). In these cases there clearly is no difference due to the independent variables, and the size of the error term is irrelevant.

The foregoing findings, together with the authors' desire to reach results relevant to the task of photo interpretation led to a second study. It was reasoned that in the training of photo-interpreters those trainees with certain scanning methods might be eliminated if in fact the methods used result in consistently poor performance. The authors decided, therefore, to rerun the five subjects who had shown the fastest detection times in the original study ( $F$  test significant at the .05 level), but because one of the subjects was not available only four were tested.

## 2. Study 2

Table 3 shows the mean detection time and scanning methods for the four subjects retested, and Table 4 shows the results of an analysis of variance.

The explanation of the significant horizontal-by-vertical-by-subjects interaction is essentially the same as in the first study. The significant horizontal-by-vertical interaction indicates that performance changes as a function of the quadrant in which the stimulus is located. The significance of this horizontal-by-vertical interaction makes doubtful the interpretation of the significant "vertical" effect. It is possible that the "vertical" effect is genuine; that is, that difficulty is a function of whether the stimulus occurs in the upper half or lower half of the matrix regardless of its left-right position. However, there is also the possibility that the difference is a reflection of the

TABLE 3  
MEAN DETECTION TIMES (IN SECONDS) AND SCANNING METHODS  
FOR SUBJECTS 2, 4, 6, AND 9

Quadrant	Subject*				Total
	2	4	6	9	
Upper left	6.4	20.8	12.0	18.8	14.5
Upper right	6.2	6.0	5.6	7.3	6.3
Lower left	5.6	9.2	17.3	17.7	12.5
Lower right	7.7	17.6	21.6	42.1	22.3

\* The method used by each subject was as follows: Subject 2—lower left, upper right, upper left, lower right; Subject 4—upper right, lower left, lower right, upper left; Subject 6—upper right, upper left, lower left, lower right; Subject 9—upper right, lower left, upper left, lower right.

TABLE 4  
ANALYSIS OF VARIANCE FOR THE DETECTION-TIME DATA OF STUDY 2

Source	df	Mean square	F
Horizontal	1	9.30	0.08
Vertical	1	772.84	6.41*
Subjects	3	601.77	4.99**
Interaction			
Horizontal-by-vertical	1	1,292.41	10.72**
Horizontal-by-subjects	3	68.27	0.57
Vertical-by-subjects	3	270.49	2.24
Horizontal-by-vertical-by-subjects	3	985.30	8.17**
Within cells	48		
Total	63		

\* Significant at the .05 level.

\*\* Significant at the .01 level.

TABLE 5  
MEAN DETECTION TIMES FOR THE FOUR QUADRANTS

Section	Left	Section	Right
Upper	14.5		6.3
Lower	12.5		22.3

horizontal-by-vertical effect. Table 5, which shows the mean detection times for the various quadrants, seems to support the interpretation that the main "vertical" effect is an artifact of the horizontal-by-vertical interaction and seems to indicate that the significance of the vertical effect is due largely to the difference in the right half of the matrix (i.e., is a function of the "horizontal" placement of stimuli, especially since the direction as well as the magnitude of the difference is reversed in the left half of the matrix).

In summary, it appears that for the individuals who are relatively proficient at the type of task under study, the position of the stimulus with respect both to horizontal and vertical orientation is a significant determinant of detection. As a next step, one should select the most efficient scanning methods and require subjects to use each of the selected methods to permit statistical evaluation of a method-by-subjects interaction. Such an evaluation would indicate whether above-average detectors excel regardless of the method used or whether different operators employ different techniques in achieving similar results.

#### D. SUMMARY

In this study, the authors investigate detection performance when the stimulus to be detected is a "C" in a 16-by-16 matrix of "Gs." Two studies



were conducted, with the subjects in the second experiment being selected on the basis of performance in the first. Results indicate that for individuals who are relatively proficient at the type of task studied, the position of the stimulus with respect to vertical and horizontal placement is an important determinant of detection time. For individuals of low proficiency, position has no effect.

*Applied Research Laboratory*

*A Division of Sylvania Electric Products, Inc.*

*Sylvania Electronic Systems*

*40 Sylvan Road*

*Waltham, Massachusetts 02154*

## AN INVESTIGATION OF COIN-SIZE ESTIMATES BY ORPHANED CHILDREN, NORMAL CHILDREN, AND EMOTIONALLY DISTURBED CHILDREN\*

*Department of Psychology, College of The Holy Cross*

---

ANTHONY F. BRONZO, JR.

---

### A. INTRODUCTION

While sensory content is always present in perception, what is perceived is also influenced by set and prior experience, so that perception is more than a passive registration of stimuli impinging on the sense organs.

The purpose of the present experiment is to show the effects of value, need, and emotion on perception. Social values and needs can modify direct impressions, such as size and shape, in the perceptual process. Klein, Schlesinger, and Meister (4) have reported that subjects overestimate a dollar sign enclosed within a circle when the circle is seen as a coin. They conclude that the increased value of the circle results in the overestimation of the dollar sign.

An experiment conducted by Bruner and Goodman (2) was concerned with need and value as organizing factors in perception. According to Bruner, there are two types of perceptual determinants: autochthonous and behavioral. Under the former are grouped those properties of the nervous system that account for phenomena like simple pair formations, closure, and contrast. Behavioral determinants control all the higher level functioning, such as perception, learning, and motivation.

Bruner further states that the human organism exists in a world of ambiguously organized sensory stimulation. What the organism sees of what is actually there perceptually represents a compromise between what is presented by autochthonous process and what is selected by behavioral processes. Such selection is determined not only by learning but also by motivational factors.

Bruner and Goodman state three hypotheses as follows:

1. The greater the social value of an object the more it will be selected through organization by behavioral determinants. . . .
2. The greater the individual need for a socially valued object, the more marked will be the operation of behavioral determinants. . . .

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on July 13, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



3. Perceptual ambiguity will facilitate the operation of behavioral determinants only in so far as ambiguity reduces the operation of autochthonous determinants without reducing the effectiveness of behavioral ones (2, pp. 33-34).

Bruner and Goodman's subjects were 30 10-year-old children of normal intelligence divided into three groups: two experimental, one control. One experimental group was called the "rich" group; the other, the "poor" group. The two experimental groups estimated the sizes of coins, while the control group estimated the sizes of gray discs. No mention of money was made to the control group.

Three of the significant results yielded by Bruner and Goodman's study were as follows: (a) socially valued objects, such as coins, were judged larger in size than were valueless gray discs; (b) the greater the value of the coin, the greater was the deviation of apparent size from actual size; and (c) the "poor" group tended to overestimate the size of the coins much more than did the "rich" group.

In 1898, Wolfe (9) performed an experiment in which school children and college students were instructed to draw circles the size of various coins and to draw rectangles the size of paper money. He concluded (a) that young children always underestimate the size of coins and bills; (b) that mature persons of intelligence overestimate the size of the silver dollar, half dollar, and quarter; and (c) that all classes of people underestimate the size of the dime, nickel, and paper money.

Myers (6) found that subjects overestimate the size of larger coins and underestimate the size of smaller coins. A replication of Bruner and Goodman's experiment was performed by Carter and Schooler (3). They secured as subjects 48 children who ranged in age from 9 to 11. Twenty-four of the children were from a settlement house and were designated as the "poor" group, while 24 children from well-to-do homes were designated as the "rich" group. The results obtained by Carter and Schooler differed from those obtained by Bruner and Goodman in that (a) the judgments of the "rich" and the "poor" children were essentially the same when the coins to be estimated were present, (b) the "poor" children overestimated the size of coins when judgments were made from memory, (c) both groups generally underestimated the size of small coins, and (d) both groups generally overestimated the size of large coins.

In 1949 Lambert, Solomon, and Watson (5) placed 54 subjects (out of 71) in an experimental group and 17 in a control group. The experimental subjects learned to turn a crank in order to receive a poker chip. When the

subject inserted the poker chip in a slot, he received automatically a piece of candy. The conjecture was that the candy would enhance the value of the poker chip, and that the increased value would result in overestimation of size (as in the coin experiment). In a pretest, both control subjects and experimental subjects overestimated the size from five to six per cent; but after 10 days of rewarded learning with poker chips, the experimental group increased its overestimation to 13 per cent (a statistically significant gain). During the same period, overestimation by the control group did not increase significantly.

In 1951 Ashley, Harper, and Runyon (1) had subjects serve as their own controls; i.e., hypnotic states were created in which subjects at one time were "rich" and at other times "poor." Results were consistent with those of Bruner and Goodman's 1947 study (both for judgments made with the coins present and for judgments made from memory).

Tajfel and Cawajsee (7) found that subjects overestimated the sizes of all coins, and that overestimation was greatest for the larger coins.

## B. HYPOTHESES

The hypotheses to be investigated in the present study are as follows:

1. that emotionally disturbed children perceive coin sizes less accurately than do normal children.
2. that orphaned (institutionalized) children perceive coin sizes more accurately than do emotionally disturbed children, but less accurately than do normal children.
3. that, of value and emotion, the latter (as seen in disturbed children) produces greater distortion than does value in the perception of coin sizes.

## C. APPARATUS AND PROCEDURE

### 1. *Apparatus*

The apparatus consisted of a rectangular wooden box ( $9'' \times 9'' \times 18''$ ) on one end of which was a five-inch, square, ground-glass screen and a knob at its lower right-hand corner. At the center of the ground-glass screen was an almost-circular patch of light cast upon the screen by a 60-watt incandescent light shining through an iris diaphragm, which could be varied in diameter from one-sixteenth of an inch to three inches (by turning the knob on the front of the box). The knob was connected to a calibrated vernier that ranged from zero to three inches.

All that was visible to the subject was the box with its ground-glass screen and the circle of light. The circle was not truly round, but contained the



familiar nine elliptoid sides found in the Bausch & Lomb iris diaphragm. It was so close to round, however, that subjects had no difficulty making the subjective comparisons required of them.

## 2. Procedure

The *Ss* were 33 males between the ages of 7 and 12 and were divided into two experimental groups and one control group. The control group consisted of 11 *Ss* from an elementary school. The first experimental group consisted of 11 *Ss* from an orphanage, and the second experimental group consisted of 11 emotionally disturbed boys from a nearby institution.

The control group lived at home and attended regular-school classes. The orphanage group lived in an institution, but attended regular-school classes. The emotionally disturbed group lived and went to school in the same institution. The *Ss* in the three groups (all from lower-class socioeconomic backgrounds) were matched for age, sex, *IQ*, parents, income, social class, religious affiliation, and length of institutionalization.

When tested, *S* sat on a chair in front of the screen on the box with the light circle slightly below his eye level. The distance between *S* and the apparatus was maintained constant throughout the experiment. The box rested on a table behind which *E* sat. *S* was instructed to make the circle of light on the box the same size as various coins. *S* was encouraged to familiarize himself with the different circle sizes that would be formed, by manipulating the diaphragm control knob.

There were two test series. First, *S* was asked to estimate from memory the sizes of coins from a penny through a half dollar in a value-ascending series. Five ascending-opening iris trials were made for each coin, and an average was computed. *S* was given no knowledge of results.

Second, using the same order of presentation, a series was conducted with the coins present. *S* examined each coin tactually for size dimensions for a period from 15 to 20 seconds; then *S* was instructed to repeat the first half of the experiment.

## D. RESULTS

The results of the present experiment are (a) coin-size estimates from memory and (b) coin-size estimates from sight. From memory, *Ss* in all three groups underestimated all coin sizes (except that for the half dollar). Emotionally disturbed *Ss* tended to be more accurate than did either of the other groups [see Table 1].

Table 1 shows, also, that the orphanage group had an average perception

TABLE 1  
COIN-SIZE ESTIMATES BY CONTROL CHILDREN, ORPHANED CHILDREN, AND EMOTIONALLY DISTURBED CHILDREN

Group	N	Penny (.44)*	Nickel (.53)	Coin Dime (.37)	Quarter (.70)	Half dollar (1.12)
Estimate from memory**						
Control children	11	.36	.45	.27	.60	1.28
Orphaned children	11	.38	.49	.25	.69	1.23
Emotionally disturbed children	11	.30	.50	.25	.72	1.56
Estimate from sight**						
Control children	11	.35	.47	.25	.59	1.06
Orphaned children	11	.37	.49	.27	.63	1.15
Emotionally disturbed children	11	.37	.49	.22	.64	1.19

\* Actual size (in inches).

\*\* In inches.



deviation closer to the correct coin sizes than did either of the other groups. For all three groups, the half dollar was the most accurately estimated coin. Table 2 shows that the control *Ss* and the orphaned *Ss* underestimated the sizes much more than overestimating them when the coins had not been seen. In this respect, control *Ss* and orphaned *Ss* underestimated the size of the coins twice as much as overestimating them. *Ss* in the emotionally disturbed group highly overestimated the sizes of the quarter and the half dollar and underestimated that of the penny, the nickel, and the dime. Emotionally disturbed *Ss* overestimated the size of the coins about three times more than did the control *Ss* and two and one-half times more than did the orphaned *Ss*.

When the coins had been seen [*cf.* Table 2], control *Ss* underestimated the sizes about three times more than they overestimated them. Orphaned *Ss* underestimated coin sizes twice as often as they overestimated them, while emotionally disturbed *Ss* remained about the same in both overestimation and underestimation (but overestimated the coin sizes three times more than did *Ss* in the control group and orphaned group).

#### E. DISCUSSION

The three groups tended to underestimate the coin sizes rather than to overestimate them. The *t* tests of the differences between coin-size estimations made by orphanage and emotionally disturbed children before and after the coins had been seen are significant for each coin except the half dollar.

After the coins had been seen, *Ss* still tended to underestimate rather than overestimate. According to Bruner and Goodman's hypothesis, the coins should have been seen as having value, and this perception in turn should influence the size perception of the coins. The results of this experiment support this hypothesis. The dime, although smaller than the penny and the nickel, has more monetary value. Emotionally disturbed *Ss*, whose social needs are thought to be the highest, overestimated the size of the dime six times more than did orphanage *Ss* ( $p < .01$ ).

The results of the present experiment show that differences exist between the coin-size estimates made by institutionalized and noninstitutionalized children. The null hypothesis that no difference exists between the coin-size estimates is rejected beyond the .01 level. Also rejected (at the .05 level) is the hypothesis that no difference exists between the coin-size estimates made by normal children and orphaned children.

#### F. SUMMARY AND CONCLUSIONS

The purpose of the present experiment was to show the effects of value, need, and emotion on perception. Thirty-three *Ss* from an elementary school,

TABLE 2  
COIN-SIZE OVERESTIMATES AND UNDERESTIMATES BY CONTROL CHILDREN,\* ORPHANED CHILDREN, AND  
EMOTIONALLY DISTURBED CHILDREN

Coin	Control children (N = 11)		Orphaned children (N = 11)		Emotionally disturbed children (N = 11)	
	Over-estimate	Under-estimate	Over-estimate	Under-estimate	Over-estimate	Under-estimate
Estimate from memory						
Penny	.06	.95	.23	.86	.32	1.88
Nickel	.03	.99	.41	.81	1.12	1.47
Dime	.13	1.23	.07	1.38	.43	1.63
Quarter	.52	1.43	.63	.93	2.17	1.94
Halfdollar	2.25	1.39	2.52	1.16	6.27	1.49
Total	2.99	6.05	3.86	5.14	10.31	8.41
Estimate from sight						
Penny	.11	1.08	.11	.91	.53	1.19
Nickel	.32	.99	.25	.78	1.50	.89
Dime	.22	1.51	.11	1.17	.19	1.74
Quarter	.32	1.55	.28	1.04	1.19	2.05
Halfdollar	1.09	1.75	1.20	.87	2.48	1.67
Total	2.06	6.88	1.95	4.77	6.02	7.54



orphanage, and a home for emotionally disturbed children were instructed to estimate from memory the size of a penny, nickel, dime, quarter, and half dollar; then to estimate the sizes after having been shown the coins. The null hypothesis that no difference exists between the coin-size estimates made by normal children and emotionally disturbed children was rejected ( $p < .01$ ); therefore:

1. Normal, orphaned, and emotionally disturbed boys tend to underestimate the sizes of pennies, nickles, dimes, and quarters before and after the coins have been seen; but the half dollar tends to be overestimated.

2. Normal boys perceive coin sizes more accurately than do emotionally disturbed boys.

3. Orphans tend to perceive coins more accurately than normal children and more accurately than emotionally disturbed children.

4. Bruner and Goodman's hypothesis on perception is substantiated.

#### REFERENCES

1. ASHLEY, W. R., HARPER, R. S., & RUNYON, D. L. The perceived size of coins in normal and hypnotically induced economic states. *Amer. J. Psychol.*, 1951, **64**, 564-572.
2. BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in perception. *J. Abn. & Soc. Psychol.*, 1947, **42**, 33-34.
3. CARTER, L. F., SCHOOLER, K. Value, need and other factors in perception. *Psychol. Rev.*, 1949, **56**, 200-207.
4. KLEIN, G. S., SCHLESINGER, H. J., MEISTER, D. E. Effect of personal values on perception. *Psychol. Rev.*, 1951, **58**, 58-96.
5. LAMBERT, W. W., & SOLOMON, R. L. Reinforcement and extinction as factors in perception. *J. Exper. Psychol.*, 1949, **39**, 637-641.
6. MYERS, G. C. A study in incidental memory. *Arch. Psychol.*, 1913, **4**. Pp. 1-108.
7. TAJFEL, H., & CAWASJEE, S. D. Value and the accentuation of judged differences. *J. Abn. & Soc. Psychol.*, 1959, **59**, 436-438.
8. WOLFE, H. K. Some judgments on the size of familiar objects. *Amer. J. Psychol.*, 1898, **9**, 137-166.

210 Ingleside Avenue  
Worcester, Massachusetts

## EXPECTANCY AND VALIDITY\*

*The Journal Press, Provincetown, Massachusetts*

LEONARD W. FERGUSON

### A. PROBLEM

Commonly, results on the value of selection tests are presented in the form of expectancy tables. An example is given in Table 1. From such a table, it is inferred that recruits with high test scores have 30 chances out of 100 for

TABLE 1  
EXAMPLE 1: PER CENT OF RECRUITS SUCCESSFUL IN  
RELATION TO TEST SCORE  
(Data are hypothetical)

Test score	Number of recruits	Recruits achieving success Number	Per cent*
High	300	90	30
Average	500	100	20
Low	200	30	15
Total	1000	220	22

\* Per cent of successes within each test-score subgroup.

achieving success, that recruits with average test scores have 20 chances out of 100 for achieving success, and that recruits with low test scores have 15 chances out of 100 for achieving success. [See Bingham (1) and Wesman (4, 5).] Further, it is inferred that chances for success on the part of high-scoring agents are one and one-half times those of recruits with average test scores (i.e., 30 divided by 20) and twice those of recruits with low test scores (i.e., 30 divided by 15).

Suppose, however, that the number of recruits were 500 rather than 1000. Then the expectancy table would look like Table 2, and one would infer that recruits with high test scores have 60 chances out of 100 for achieving success, that recruits with average test scores have 40 chances out of 100 for achieving success, and that recruits with low test scores have 30 chances out of 100 for achieving success. In each instance, the chances of success are double those in the first example.

\* Received in the Editorial Office, Provincetown, Massachusetts, on July 16, 1964, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



TABLE 2  
EXAMPLE 2: PER CENT OF RECRUITS SUCCESSFUL IN  
RELATION TO TEST SCORE  
(Data are hypothetical)

Test score	Number of recruits	Recruits achieving success	
		Number	Per cent*
High	150	90	60
Average	250	100	40
Low	100	30	30
Total	500	220	44

\* Per cent of successes within each test-score subgroup.

Does the set of odds just given mean that the test involved in Table 2 is more *predictive* than the test involved in Table 1?

By no means!

In both examples, the ratios among the per cents are identical. The expectancies differ not because of a difference in test predictiveness but because of a difference in the per cents or proportions of *all* recruits achieving success: 22 per cent in the first example, 44 per cent in the second.

## B. SOLUTION

The foregoing effect on expectancies—of change in the per cent or proportion of successes—has long been known. In fact, it was just such knowledge that led to the Taylor-Russell tables (3). If one knows (*a*) the validity of a test, (*b*) the per cent or proportion of past recruits successful, and (*c*) the per cent or proportion of applicants to be hired, one can estimate—with the aid of the Taylor-Russell tables—(*d*) the probable per cent or proportion of successes to be obtained by the employment of recruits having scores above some specified critical level. Use of the Taylor-Russell tables requires, however, knowledge of the validity of a test in terms of a correlation coefficient expressing the degree of relation between a criterion and its test predictor. Unfortunately, this correlation often is not computed; other times it cannot be computed; and, in many instances, it should not be computed because assumptions underlying its application and interpretation are not met.

There is needed, therefore (for reporting test-validity results in expectancy-table style), some *standardized* method that permits comparison of the relative *predictive* powers of two or more tests even though correlation coefficients are not available and even though per cents or proportions of successes differ among the total recruit groups within which test-score subgroups are to be compared.

To accomplish the end just indicated, the data in Tables 1 and 2 can be

TABLE 3  
REARRANGEMENT OF DATA FROM TABLES 1 AND 2

Test score	Recruits		Successes	
	Number	Per cent*	Number	Per cent**
From Table 1				
High	300	30.0	90	40.9
Average	500	50.0	100	45.5
Low	200	20.0	30	13.6
Total	1000	100.0	220	100.0
From Table 2				
High	150*	30.0	90	40.9
Average	250	50.0	100	45.5
Low	100	20.0	30	13.6
Total	500	100.0	220	100.0

\* Per cent distribution of test scores within total recruit group.

\*\* Per cent of distribution of test scores within success group.

rearranged as in Table 3, after suggestion derived from Strong (2, pp. 39-44). Data arranged as in Table 3 eliminates the effect of differences in per cents or proportions of total recruit groups considered successful. In such arrangement it is clear that high, average, and low test scores are present in the same per cent or proportion in both recruit groups.

To determine expectancies for success, the two per cents in each row of Table 3 need to be added to produce the totals presented in Column 1 of Table 4. Next, each of these totals is divided into the appropriate per cent

TABLE 4  
EXPECTANCIES BASED ON EQUALIZED CRITERION GROUPS AND UPON  
EQUALIZED TEST-SCORE GROUPS

Test score	Sum of per cents in Columns 2 and 4 of Table 3	Per cent in Column 4 of Table 3 divided by sum in Column 1 of Table 4
First example		
High	70.9 (i.e., 30 + 40.9)	58 (i.e., $40.9 \div 70.9$ )
Average	95.5 (i.e., 50 + 45.5)	48 (i.e., $45.5 \div 95.5$ )
Low	36.6 (i.e., 20 + 13.6)	37 (i.e., $13.6 \div 36.6$ )
Second example		
High	70.9 (i.e., 30 + 40.9)	58 (i.e., $40.9 \div 70.9$ )
Average	95.5 (i.e., 50 + 45.5)	48 (i.e., $45.5 \div 95.5$ )
Low	36.6 (i.e., 20 + 13.6)	37 (i.e., $13.6 \div 36.6$ )

in Column 4 of Table 3 to produce the ratio given in Column 2 of Table 4. Table 4 shows that both sets of expectancies are identical, as they should be. The expectancies show the tests to be equally predictive *regardless* of the per cents or proportions of total recruits successful.



## C. CONCLUSION

As the foregoing discussion illustrates, for the proper determination of test validity in a predictive sense equalized *criterion* groups as well as equalized *test-score* groups are needed. Table 4 gives results based on equalized criterion groups as well as on equalized test-score groups.

From the data in Table 4, one can infer as a measure of predictive validity that recruits with high test scores have 58 chances for success and 42 chances for failure, that recruits with average test scores have 48 chances for success and 52 chances for failure, and that recruits with low test scores have 37 chances for success and 63 chances for failure. Regardless of which recruit sample is taken the chances (or odds) remain the same.

At this point someone is sure to ask, "Does not the foregoing result and interpretation overlook the fact that in the situation shown in Table 1 a recruit with a high test score *really* has only 30 chances out of 100 for achieving success and that in the situation shown in Table 2 a recruit *really* has 60 chances out of 100 for achieving success?"

No! The chances just cited differ *only* because of the difference in the per cents or proportions of recruits successful—not because of variation in test-score validity.

In the practical employment situation, one assuredly must pay attention to the per cent or proportion of total recruits successful—this is the *raison d'être* of the Taylor-Russell tables; but one should not give this attention *at the expense* of data needed to determine test predictiveness; i.e., one should not give attention to the per cent or proportion of total proportion of total recruits successful in a way that *misleads* when the issue is one of comparing the predictive powers of two or more tests used in situations in which the per cents or proportions of recruits successful is likely to vary.

## REFERENCES

1. BINGHAM, W. V. Expectancies. *Educ. & Psychol. Meas.*, 1953, **13**, 47-53.
2. STRONG, E. K., JR. Vocational Interests 18 Years After College. Minneapolis: Univ. Minnesota Press, 1955.
3. TAYLOR, H. C., & RUSSELL, J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *J. Appl. Psychol.*, 1939, **23**, 565-578.
4. WESMAN, A. G. Expectancy tables, a way of interpreting test validity. *Test Service Bulletin No. 38*, Psychological Corporation, New York, 1949.
5. ———. Better than chance. *Test Service Bulletin No. 45*, Psychological Corporation, New York, 1953.

*The Journal Press*

2 Commercial Street

Provincetown, Massachusetts

## ESP: MOTIVATION AS A FACTOR IN ABILITY\*

*Department of Psychology, American International College*

---

RICHARD C. SPRINTHALL AND BARRY S. LUBETKIN

---

### A. INTRODUCTION

ESP or "extrasensory perception," a term coined by Dr. J. B. Rhine, means the acquiring of information through channels other than the known physiological sense receptors. Variouslly called "mental telepathy," "mind reading," and "mental radio;" ESP has been an object of man's fascination for centuries, although it is only in relatively recent years that attempts have been made to subject this phenomenon to the formal scrutiny of the scientific method. Since 1930 serious attempts have been made, especially by Dr. J. B. Rhine in the United States and later by Dr. S. G. Soal in England, to establish ESP as a reliable scientific fact. Rhine feels that ESP became an established fact soon after his own quest began, for he says:

Once methods of standardizing the testing of ESP had been developed, the logical next step was to design an experiment that would be crucial and so take the question of the occurrence of psi out of the realm of debatability. For the Duke investigators this milestone was reached in 1933 when the Pearce-Pratt series of experiments was completed (4, p. 6).

This "crucial" experiment, with J. G. Pratt (then a graduate student under Rhine) as sender and Hubert Pearce (a divinity-school student) as receiver, utilized the Zener or ESP cards: a deck of 25 cards with five different "suits" of five cards each. These suits are composed of various geometric symbols: a square, a circle, a star, wavy lines, and a cross; thus, chance would predict that on any given run of 25 cards, there should be five correct calls. Pearce, in the 1933 study, made 119 hits out of 300 calls; while chance would predict only 20 per cent or 60 hits. Says Rhine, "Such a result could hardly be thought of as explainable by chance," and, "Nothing, then, but ESP could explain the results" (4, p. 7).

Even now, this is the basic procedure used by Rhine (that is, to run a large number of trials), and when a subject scores at a rate significantly higher than 20 per cent, Rhine infers ESP. H. L. Mencken writes "In plain

---

\* Received in the Editorial Office, Provincetown, Massachusetts, on July 15, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.



language, Professor Rhine segregates all those persons who, in guessing the cards, enjoy noteworthy runs of luck, and then adduces those noteworthy runs of luck as proof that they must possess mysterious powers" (3). Clarence Leuba, in an appraisal of Rhine's work, also makes the same point (2, p. 221).

Edwin G. Boring also criticizes the Rhine probability model: "a model whose fit can never be tested empirically because every observed deviation from expectation still fits the model which tells you merely how improbable the deviation was" (1, p. 112).

In a good experiment you would turn it off—the control experiment—and note the number. If the difference were large enough to show that you are probably not in the two series dealing with the same population of guesses, then you have ESP and also an indication of how surely you have it (1, p. 112).

"But how," asks Boring, "do you turn telepathy on or off, controlling the independent variable?" (1, p. 112)

Since Boring's question, Rhine has suggested an idea whereby this technique might be applied. In 1964 Rhine made a strong case for subject motivation as being the crucial ingredient in ESP success. Although this thought had occurred to Rhine before, and is mentioned in all his books, his 1964 article presents the case in its boldest fashion.

In general, subject motivation to score high has long stood out as the mental variable that seems most closely related to the amount of psi effect shown in test results (5, p. 43).

Rhine then goes on to review seven cases of exceptionally high scoring ability over the past 35 years and notes the connection between these cases and a high motivation level. The method of inducing motivation ranges from betting one hundred dollars that a subject can not correctly call a card (in the case of Hubert Pearce) to promising 50 cents for a perfect score of 25 (in the case of Lillian) to offering a new set of electric trains for a score of 20 (in the case of Franklin). This technique created fantastically high scores in all seven cases reported by Rhine (5, p. 47).

## B. THE PROBLEM

Since the establishment of ESP rests on a probability model and not on experimental evidence, and since Rhine sees motivation as possibly the most powerful ingredient in ESP success, an experiment was designed to test this hypothesis. The design allowed for the testing of the motivation hypothesis

by active manipulation of a stimulus variable and checking for reliable response differences. Also built into the design was an opportunity to re-check a conclusion by Sprinthall (6), who found (contrary to Rhine's hypothesis) that the attitude of a subject toward ESP is not related to that subject's ESP ability (6, p. 69).

### C. PROCEDURE

Fifty subjects were selected from among two classes at American International College. All subjects were volunteers and were told only that the experience would prove to be extremely interesting. The subjects were then given the "Attitude-Toward-ESP" Test, a Likert-type scale of 18 items in which the subjects select for each item (among five alternatives) from "strongly agree" to "strongly disagree" (6, p. 67).

Next, the subjects were randomly assigned to the experimental (motivation) group and the control (no induced-motivation) group. Each subject was then brought individually, at 10-minute intervals, into the psychology laboratory for the ESP-ability test. The testing was done over a two-day period (the control subjects tested the first day, and the experimental subjects the second). The reason for this was to ensure that control subjects would not be negatively affected by having the news leaked to them that a reward was being offered to members of the other group. By the time the first experimental subject was made aware of the offer of financial reward, all the control subjects had been tested.

On being ushered into the experimental room, each subject was requested to sit down and make himself comfortable (removing coat, smoking a cigarette, etc.), and the following instructions were given to all subjects:

Good afternoon. I'd like to thank you for being here and taking an interest in our project. We're concerned in this study with determining and obtaining information concerning extrasensory perception, ESP, or as it is sometimes called, mental telepathy. This is still a wide-open field, so we feel that anything in the way of new knowledge which we here at A.I.C. can contribute will be more than worth while. On the sheets, which you see before you, are five different symbols: a cross, wavy lines, a star, a square, and a circle. In the next few minutes I will attempt to convey these symbols to you, using this deck of cards. On each of the 25 cards there is a symbol such as those on the papers before you, here—you may see for yourself. What I will do will be to look at each card, individually, and concentrate upon sending the symbol across to you. You, in turn, will merely call out the impression that you have received. I shall record it, and we'll proceed to the next card. Do you have any questions?



Then to the 25 subjects in the experimental group the following instructions were added:

As an added incentive to you, for assisting us, and in order to discover the exceptionally good subjects, the psychology department here at A.I.C. has made the following offer, in all seriousness and in good faith: we will present one hundred dollars in cash to anyone in this study who can call 20 out of 25 of these cards correctly. Let me repeat. This offer is being made in total earnest. You will be awarded one hundred dollars if you can name 20 out of 25 cards or better correctly. Your scores will be given to you immediately after we run through the cards.

While listening to the instructions, the subject found placed on the desk in front of him five separate sheets of white paper on which were drawn the five ESP symbols. This was done to allow each subject time to familiarize himself with the various symbols and, to ensure that they were remembered, they were kept in place throughout the ESP test.

The experimenter, always the same person, a male whose attitudes toward ESP were at least open-minded, if not pro, then placed the already shuffled deck on top of the table. Four shuffles were used for each new subject. The experimenter then took one card from the top of the deck, holding it so that only the back of the card was visible to the subject. He then asked for the subject's response, first, by saying "begin"; then, for the rest of the cards, by saying "next." After each response the used card was placed face down and out of the direct view of the subject.

Often, a subject would ask how long he could concentrate before giving a response or if he should close his eyes. It was always pointed out that he could take as long as he wished and that he could assume any pose that would best allow him to concentrate, eyes open or closed.

Immediately after the 25th response, the experimenter checked each response on his ESP score pad. The subject was allowed to watch this procedure; so, if a member of the experimental group, he would have no conflict regarding the possibility of winning the one hundred dollar prize. After hearing the result, the subject was thanked and requested that he not discuss his experience for the next few days.

#### D. RESULTS

Nobody won the one hundred dollars, the highest number of hits recorded being 10. On the ESP-ability test the mean number of hits of the 25 control subjects was 5.56, with an *SD* of 2.06. In the experimental (motivated)

group, the mean was 5.40 (less than that of the control group), with an *SD* of 2.10.

An independent *t* was computed and was found to be .27. With alpha set at .05, and with 48 degrees of freedom, the null hypothesis is accepted, and the authors conclude that the independent variable had no effect; i.e., motivation did not affect performance on the ESP-ability test.

The mean score on the Attitude-toward-ESP Test was 54.06, with an *SD* of 9.88. The range was 31 to 74. Because the theoretical range on the test is 18 to 90 (with a theoretical mean of 54), our subjects were about evenly divided as to their attitudes toward ESP. In Sprinthall's 1964 study, the mean attitude score on the same test (for 146 subjects) was 58.95, with an *SD* of 9.15 (6, p. 68).

As was the case in the previous study, the subject's attitude scores were compared to their ability scores; again the computed *r* (.094) was found to be not significantly different from zero. (In Sprinthall's study, it was .016.) Therefore, one's attitude toward ESP is independent of one's success in correctly naming the ESP symbols.

#### E. SUMMARY AND CONCLUSIONS

1. An attempt was made to discover if, as Rhine contends, motivation is a major factor in influencing one's scores on an ESP-ability test. This attempt was made via the experimental method, with active stimulus manipulation.

2. Two groups of 25 subjects each were selected and given the ESP-ability test. The experimental (motivated) group was told a one hundred dollar prize would be given to anyone having a score of 20 or more correct responses.

3. An independent *t* was computed and found to equal .27, showing no significant difference between the groups.

4. All the subjects took an Attitude-toward-ESP Test, and a Pearson *r* correlating these scores with ability scores yielded a correlation coefficient of .094, not significantly different from zero.

5. Motivation did not affect, and attitudes toward ESP did not relate, to one's success in guessing the symbols on a deck of ESP cards.

#### REFERENCES

1. BORING, E. G. The Present Status of Telepathy. *Amer. Scientist*, 1955, **43**, 108-117.
2. LEUBA, C. An experiment to test the role of chance in ESP research. *J. Parapsychol.*, 1938, **2**, 217-221.



3. MENCKEN, H. L. Every man his own radio. *Baltimore Evening Sun*, Baltimore, Maryland, 1937, 56, 19.
4. RHINE, J. B. A Brief Introduction to Parapsychology. Durham, N.C.: Duke Univ. Press, 1956.
5. ———. Special motivation in some exceptional ESP performances. *J. Parapsychol.*, 1964, 28, 42-50.
6. SPRINTHALL, R. C. ESP: Some attitudinal factors relating to ability. *J. of Psychol.*, 1964, 57, 65-69.

*Department of Psychology*  
*American International College*  
*Springfield, Massachusetts 01109*

